2008

Signature Identification in the Light of Science and Experience

Roger C. Park
UC Hastings College of the Law, parkr@uchastings.edu

Signature Identification in the Light of Science and Experience

ROGER C. PARK*

INTRODUCTION

The *Daubert* case encourages judges to ask whether forensic identification expertise is valid, not merely whether it is accepted among practitioners.[1] The example of DNA has shown what real science can do, and has highlighted the shortcomings of other forms of forensic science.[2] The combined effect of *Daubert* and DNA has contributed to skepticism about forensic identification techniques. This skepticism may lead to exclusion of evidence or to procedural limits aimed at making the expertise more trustworthy or preventing it from having undue weight. I will discuss these two alternatives in the context of the specific forensic problem of signature authentication expertise of forensic document examiners (FDEs), after first considering general principles applicable to experience-based expertise.

I. EXPERIENCE-BASED EXPERTISE: GENERAL CONSIDERATION

Experience counts. It brought important advances long before humans devised the alphabet, much less the scientific method.[3] In writing about *Daubert*, academic lawyers should be wary of making too broad a

---

* James Edgar Hervey Distinguished Professor of Law, University of California, Hastings College of the Law. I wish to thank David Faigman, Ed Imwinkelried, Michael Risinger, and Michael Saks for their helpful comments on earlier drafts of this Article, while exonerating them for any responsibility for its contents. The source-finding work of Charles Marcus of the U.C. Hastings library faculty was invaluable. Billy Minshall and Lila Mirrashidi of the U.C. Hastings Class of 2009 contributed able. research assistance, and Beverly Taylor has my thanks for her accurate and timely administrative work.

1. Daubert v. Merrill Dow Pharm., Inc., 509 U.S. 579, 593 (1993). *See also* David L. Faigman, *Is Science Different for Lawyers?*, 297 SCIENCE 339, 339–40 (2002); David L. Faigman, *Making the Law Safe for Science: A Proposed Rule for the Admission of Expret Testimony*, 35 WASHBURN L.J. 401 (1996).

2. *See* Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 SCIENCE 892, 893 (2005).

3. For interesting accounts of human progress in technology and agriculture in prehistoric times, see JARED DIAMOND, GUNS, GERMS, AND STEEL: THE FATES OF HUMAN SOCIETIES (2005), and NICHOLAS WADE, BEFORE THE DAWN: RECOVERING THE LOST HISTORY OF OUR ANCESTORS (2006).

denunciation of experience-based expertise.[4] Unsystematic inductions from experience are often useful, even in deciding how to use scientific studies.[5] Of course, it is indisputable that experience can lead us astray, and flaws in human reasoning distort our interpretation of it.

The question for judges who must screen expert testimony is not whether experience is ever helpful. At heart, it is the familiar evidence problem of assessing probative value against the counterweights of cost, prejudice, and waste of time, while bearing in mind alternative forms of proof.[6] The legal literature has provided some helpful guidelines.[7]

One important question is whether experts who offer experience-based expertise are in a position to learn from their experience. Judge McKenna's well-known example of experience-based expertise, the testimony of a harbor pilot, presents us with a case where the answer clearly seems to be "yes."[8] The harbor pilot learns from experience whether his beliefs about how to bring a ship safely to its berth are valid.[9] There is a "feedback loop" and a penalty for being wrong.[10] The pilot is probably in a better position to learn from experience than the documents examiner, because the documents examiner often does not get accurate feedback about whether her experience-based inductions were correct. The fact that a jury reached a verdict consistent with her

4. I am using the term "experience" because it is familiar. I intend to exclude expertise that is based on personal experience, on the reported experience of others, and on introspection. Other terms that might describe what I am talking about are "fireside inductions" or "unsystematic inductions."

5. Fireside inductions are needed when we make judgments about the internal and external validity of experiments. For example, common-sense inductions tell us that handwriting experts might try harder on proficiency tests than lay subjects, so that differences in performance could be due to effort instead of expertise; and that studies showing superior expert performance on signature authentication might not generalize to other tasks, such as attribution of authorship in cases of disguised hand printing. D. Michael Risinger, *Handwriting Identification*, *in* 4 MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY 379, § 33:14 (David L. Faigman et al. eds., 2007–2008 ed.) [hereinafter MODERN SCIENTIFIC EVIDENCE].

6. *Cf.* Old Chief v. United States, 519 U.S. 172, 180, 182–84 (1997) (describing an analogous process under Rule 403 of weighing prejudice and other costs against probative value while considering alternative means of proof).

7. *See* authorities cited *infra* notes 38–47.

8. *See* United States v. Starzecpyzel, 880 F. Supp. 1027, 1041, 1046 (S.D.N.Y. 1995).

9. This is a supposition on my part. It seems plausible that harbor pilots would get negative feedback upon going aground or damaging the ship. For an example, see Henry K. Lee & Carl Nolte, *Cosco Busan Pilot Charged with Pair of Crimes*, S.F. CHRONICLE, Mar. 18, 2008, at B1 (reporting that harbor pilot who crashed into Bay Bridge was indicted for misdemeanors based on negligent conduct, including sailing in the fog, speeding, failing to check navigation equipment, and not using radar). However, they may make other mistakes that go undetected or unpunished. For example, in the absence of systematic record-keeping and controlled testing, experience-based experts can have mistaken beliefs that go undetected because they take the form of unnecessary precautions. For examples of unnecessary precautions, see Jack Foisie, *U.S. Pilots Fly Modern Jets but Cling to Superstitions*, 30 W. FOLKLORE 140 (1971).

10. Mark P. Denbeaux & D. Michael Risinger, Kumho Tire *and Expert Reliability: How the Question You Ask Gives the Answer You Get*, 34 SETON HALL L. REV. 15, 56 (2003).

conclusions does not tell her much.[11]

In the absence of a feedback loop and penalty, it is reasonable for courts to put the burden of doing research validating the expertise upon its proponents, rather than upon the doubters. By doing so, courts would merely be requiring that the expertise be tested. Sometimes experience itself is enough of a test, and the expert can claim to have learned in the "school of hard knocks." But if the expert has no trustworthy way of determining whether he was wrong, there has been no test.

Handwriting identification doubters have been criticized for mounting a basically negative attack, as opposed to providing their own empirical studies showing that the expertise is untrustworthy.[12] But where the expertise is not tested in ordinary practice because of the lack of a feedback loop and penalty, it is reasonable to require that it be tested in some other way.

One way of validating handwriting identification expertise would be to test its premises scientifically, in which case the expertise would no longer be experience-based in the sense I am using the term here. Experience would have been used to generate hypotheses, but the resulting conclusions would be based on scientific testing. Experience-based expertise can also be tested by the "black box" method: instead of testing underlying premises and cause-and-effect hypotheses, the researcher tests the proficiency of the expert at doing what the expert claims to be able to do.[13] A controlled experiment testing whether forgeries have blunt endings more often than genuine signatures is an example of the former approach,[14] while a proficiency test that evaluates

---

11. *Id.*; *see also* Randolph N. Jonakait, *The Assessment of Expertise: Transcending Construction*, 37 SANTA CLARA L. REV. 301, 342–43 (1997).

12. Andre A. Moenssens, Meeting the *Daubert* Challenge to Handwriting Evidence: Preparing for a *Daubert* Hearing, http://forensic-evidence.com/site/ID/ID_FBI.html (last visited Apr. 20, 2008).

> When the subject of testing and validity comes up, it should also be pointed out that no research has ever surfaced that denies the existence of the skill of competent handwriting examiners or that proves that such skill does not exist! In other words, the only critical publications are the Risinger-Denbeaux-Saks articles, which do not deny explicitly the existence of the skill, but state only that they have not been convinced the skills exists. Their disbelief does not constitute proof of the non-existence of the skill of handwriting examiners. There are no studies showing that the skill of competent forensic document examiners in identifying authors of handwritings does not exist. The critics have it backwards.

Id.

13. *See infra* Part III.

14. *See* David Black et al., *The Frequency of the Occurrence of Handwriting Performance Features Used to Predict Whether Questioned Signatures Are Simulated*, 15 J. FORENSIC DOCUMENT EXAMINATION 17 (2003). The investigators had thirty-one participants create 620 simulated signatures. *Id.* These were compared to 177 genuine signatures produced by one participant. *Id.* An FDE was asked to examine the genuine and simulated signatures and count blunt endings. *Id.* at 22. The FDE found blunt endings in 16.9% of the genuine signatures and 72.7% of the simulated signatures. *Id.* at 23 tbl. 1. The value of the study is diminished by the fact that a single subject created all of the genuine signatures; it is possible that her genuine signature was unusually devoid of blunt endings.

whether document examiners are more accurate than lay people in detecting forgeries is an example of the latter approach.[15] On the specific task that is examined in this Article, signature authentication, there has been a significant amount of proficiency testing, and this testing will be discussed later in this Article.

Another question is whether experience-based testimony will be prejudicial because jurors think it is scientific when it is not. If the specialist clothes herself in the language of science when the testimony is not really the product of the scientific method, it may be given too much weight.[16] The expert who assumes the aura of science while really basing her testimony on unsystematic inductions creates the worst of both worlds, by appropriating the prestige of science without its acknowledgment of uncertainty; delivering testimony that may be more clear and definite (and hence convincing) than what a real scientist would deliver.[17]

When the expert witness relies upon unrecorded personal experience, the testimony can also be prejudicial because the witness can use the cloak of experience to conceal bias or ignorance. Personal experience is, after all, personal. For example, an expert's claim that "I have never seen a similar instance" may mean, "You don't know what I have seen and what I haven't, so I can say this and get away with it."[18] Testimony that an opinion is based upon "my 26 years of experience in the field" may mean, "It's really a surmise on my part. I believe it to be true, but I can't really tell you why I think that. It's really more of an impression that I have than anything else but I can't say that it's a surmise or a vague impression, could I?"[19]

Because experts who rely upon personal experience have leeway in tailoring their testimony, safeguards against bias are important. When the crime lab is an arm of the prosecution and there is no attempt to wall off the expert from extraneous information indicating that the defendant

---

15. Michael J. Saks, *Forensic Identification: From a Faith-Based "Science" to a Scientific Science*, FORENSIC SCI. INT'L (forthcoming 2008).

16. *See* United States v. Starzecpyzel, 880 F. Supp. 1027, 1048 (S.D.N.Y. 1995) ("With regard to scientific experts, a major rationale for *Frye*, and now *Daubert*, is that scientific testimony may carry an 'aura of infallibility.' . . . Skilled experts generally present less of a problem, as, with all due respect, accountants are unlikely bearers of an aura of infallibility." (citation omitted)).

17. *See* Jonakait, *supra* note 11, at 308 (citing Anthony Champagne et al., *An Empirical Examination of the Use of Expert Witnesses in American Courts*, 31 JURIMETRICS J. 375, 380 (1991) (proposing that "jurors more willingly accept experts who present information nontechnically and give firm conclusions than those who do not")).

18. For this "translation," see John I. Thornton & Joseph L. Peterson, *The General Assumptions and Rationale of Forensic Identification*, *in* MODERN SCIENTIFIC EVIDENCE, *supra* note 5, § 29:21. *Cf.* Jonakait, *supra* note 11, at 310 & n.30 ("[T]he fact-finder need never take a scientific expert witness' 'word for it.'" (quoting MICHAEL J. SAKS & RICHARD VAN DUIZEND, THE USE OF SCIENTIFIC EVIDENCE IN LITIGATION 5 (1983)) (alteration in original)).

19. Thornton & Peterson, *supra* note 18.

is guilty, the identification expert is especially susceptible to bias.[20] For example, in a terrorism case, the FBI fingerprint identification experts who erroneously identified the fingerprints of an Oregon lawyer as matching a latent print found on a bag of detonators in Madrid may have been influenced by knowledge that the lawyer had converted to Islam and had represented a Taliban sympathizer.[21] So the presence of safeguards against bias is another factor the judge could take into account in deciding whether to exclude or limit the testimony.

Another consideration is the distinction between descriptive expert opinions and evaluative ones, a distinction that turns upon the degree of inference and speculation involved in an experience-based opinion. It is easy to make the case for using an experience-based expert for purposes of educating the jury about occurrences and conditions observed by the expert, such as practices within an industry. The "summarizational" expert who, for example, testifies to trade practices differs from the lay witness only in that he is allowed to be less concrete; to summarize instead of telling a series of anecdotes.[22] It is more difficult to justify testimony by the experience-based "translational" or interpretive expert (the expert who says what something means).[23] The translational witness may be wrong in her inferences in ways that are misleading and hard to penetrate; the process of getting from sense data to testimony is complicated and extraordinary.

One can imagine handwriting expertise that is purely descriptive and summarizational. This might include testimony about class characteristics, such as styles taught in school, that unwary jurors might mistake for individual characteristics. For example, when I was in grade school, I was taught to make sevens with a hook. I do not remember seeing a hookless, crossed seven until after I went to a different part of the country to go to college. A jury that thinks a crossed seven is an unusual quirk, like a heart used in place of a period, would obviously profit from learning that many people write it that way.[24]

---

20. D. Michael Risinger et al., *The* Daubert/Kumho *Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion*, 90 CAL. L. REV. 1, 21 (2002).

21. Jennifer L. Mnookin, *The Achilles' Heel of Fingerprints*, WASH. POST, May 29, 2004, at A27; *see also* Robert B. Stacey, *Report on the Erroneous Fingerprint Individualization in the Madrid Train Bombing Case*, 7 FORENSIC SCI. COMM. (2005), http://www.fbi.gov/hq/lab/fsc/backissu/jan2005/special_report/2005_special_report.htm. Mr. Stacey, Unit Chief of the FBI Quality Assurance and Training Unit, concluded that context effect influenced the initial identification (though he did not specifically state that the examiners were aware of Mr. Mayfield's Muslim connections), and that subsequent identifications were "tainted" because, among other things, the subsequent examiners were aware of the previous examiner's conclusion and "[t]o disagree was not an expected response." *Id.*

22. D. Michael Risinger, *Functional Taxonomy of Expertise*, *in* 1 MODERN SCIENTIFIC EVIDENCE, *supra* note 5, § 2:3.

23. *Id.*

24. *See generally* Marvin L. Simner et al., *A Comparison of the Arabic Numerals One Through Nine, Written by Adults from Native English-Speaking vs. Non-Native English Speaking Countries*, 15

To use an example from case law, consider the testimony of the defense expert in the *Fujii* case.[25] There, the prosecution offered the testimony of a forensic document examiner who purported to be able to tell that the hand printing on address labels on drug shipments was Mr. Fujii's hand printing.[26] This was an adventurous claim, for a number of reasons, including poor performance of experts on proficiency tests involving hand printing,[27] and the difficulty of assigning authorship to intentionally disguised writing.[28] (The printer of the labels and, if guilty, the defendant who is asked to supply exemplars both would have had plenty of motivation to disguise their printing.) Moreover, Mr. Fujii's hand printing may have been particularly nondistinctive. In conjunction with the *Daubert* hearing on the admissibility of FDE testimony, the defendant submitted an affidavit of Mark Litwicki, the Director of Loyola University's English as a Second Language program.[29] As described by the court, at the *Daubert* hearing on the admissibility of the expert testimony, the Court described the defendant's affidavit as follows:

> Mr. Litwicki avers that he has had extensive experience examining the handwriting styles of foreign students, including Japanese students, as he has taught many Japanese students in this country. He further avers that he spent two years teaching English to Japanese students in Japan and is "especially familiar with the manner in which Japanese-trained writers make the characters of the English alphabet." Mr. Litwicki avers that the Japanese language requires its students to spend a great deal of time learning to write several thousand Japanese characters, that uniformity of characters is "an important and valued principle of Japanese handwriting," that Japanese students "spend many years attempting to maximize the uniformity of their writing" and that "Japanese-trained writers also tend to write English characters in a very uniform manner." Mr. Litwicki concludes, "In my opinion, it would be very difficult for an individual not familiar with the English handwriting of Japanese writers to identify the subtle dissimilarities in

---

J. FORENSIC DOCUMENT EXAMINATION 1 (2003) (offering an example of a study providing information about this sort of class characteristic). The Simner et al. study found, among other things, that among writers of Arabic numbers, 97% of the 113 German university students surveyed wrote crossed sevens sometimes also adding a hook, whereas 98% of the 86 Japanese students surveyed wrote hooked sevens without a cross. *Id.* at 14 tbl.7.

25. United States v. Fujii, 152 F. Supp. 2d 939, 941–42 (N.D. Ill. 2000).

26. *Id.* at 939–40.

27. *Id.* at 941 ("Michael Saks, who testified for the defense, testified that he was aware of only one study of the reliability of handprinting identification, and in that study, only 13% of the handwriting experts tested got the right answer; 45% identified the wrong person."); Risinger, *supra* note 5, at 513 (describing the hand printing proficiency test). It seems to have been a particularly hard test, in which a professional document examiner wrote a mock holdup note simulating the hand printing of another subject, and 45% of the subjects assigned authorship to the forger instead of to the actual author. *See id.*

28. Risinger, *supra* note 5, at 505 (citing ALBERT S. OSBORN, QUESTIONED DOCUMENTS 13–14 (1910) [hereinafter OSBORN, 1910 edition]).

29. *Fujii*, 152 F. Supp. 2d at 941.

the handwriting of individual writers."[30]

Now, suppose that the *Fujii* case had gone to trial, without the FDE expert. The address labels of disputed authorship are an exhibit, as are the demand exemplars showing known examples of Mr. Fujii's hand printing. The jury would be given the daunting task of making comparisons. Surely the jury would be entitled to hear the testimony of Mr. Litwicki about how Japanese writers of English make very uniform characters. Except for its conclusion, the affidavit is descriptive and summarizational.

I will put aside descriptive testimony like that of Mr. Litwicki for purposes of this Article. The testimony I will be evaluating is testimony similar to that contained in Appendices 1 and 2, where there is a good measure of expert inference and interpretation as well as description.

Another consideration in evaluating experience-based expertise is what, with some misgivings, I will call inherent plausibility. By this I mean that the judge making the admissibility determination is entitled to use her own experience, introspection and fireside inductions in deciding whether the theories of the expert are persuasive. It is not plausible to think that the gravitational attraction of the stars at the moment of birth affects personality, when the gravitational attraction of the attending physician would be greater; however, it is plausible to believe that someone tracing someone else's signature might write more slowly than someone writing her own signature. Inherent plausibility was an important element in *Kumho Tire Co. v. Carmichael*,[31] where it was hard to believe that the expert could formulate his complicated, nuanced criteria for detecting tire abuse with accuracy in the absence of any empirical testing, or that he could be accurate in his complicated, hard-to-disprove inferences about what would cause a worn tire to blow out, while lacking the ability to make ball-park estimates about a simple fact that could have been disproved had the expert been mistaken (the number of miles of wear that the tire had endured).[32] In assessing inherent plausibility, the judge is entitled to consider not only whether she herself finds the theory plausible, but whether other people she respects finds it plausible—in other words, to use the authority of others as one guide to decision.

Finally, the judge should consider the alternatives to using the specific type of experience-based testimony. One form of considering the alternatives is to consider alternatives that exist at the time of trial. For example, testimony about the cause of an illness based on clinical experience might be excluded if more trustworthy epidemiological

30. *Id.*

31. 526 U.S. 137, 141, 153, 157–58 (1999).

32. *See id.* at 137.

evidence is available.[33] A more adventurous approach is to ask what alternatives are conceivably possible—in other words, to be demanding consumers and to take into consideration not only whether better research is actually available, but whether the field would be encouraged to do better research if judges excluded experience-based expertise.[34]

If this factor is taken into consideration, then forensic identification may be a better target for a demanding consumer than experience-based behavioral expertise, because true experiments are easier to conduct in the former area. It is feasible to test propositions about handwriting with randomized trials, but not propositions about coerced confessions or rape trauma, because one cannot assign subjects to undergo coercive treatment.[35] These obstacles do not stand in the way of experiments examining premises about handwriting, or the proficiency of handwriting experts, though the fact that so far there is not much overlap with established academic inquiry may be a formidable obstacle to finding academic researchers who are motivated to do the studies.

One can argue for a generally permissive view toward forensic identification testimony by pointing out that it is probably better than eyewitness identification testimony, and that overly broad exclusion of expert testimony could lead to too much reliance on dubious eyewitness identification testimony.[36] But this argument can never be a justification for admitting expert testimony that is affirmatively misleading, though perhaps it might be an argument that marginally helpful expertise is worth the time and money. But when the danger of being misled by eyewitness testimony is important, a better remedy might be to allow testimony by experts about flaws in eyewitness testimony, rather than allowing dubious expertise about other forms of identification.[37] At any rate, trying to take the danger of prejudice from flawed eyewitness testimony into account on a case-by-case basis would diminish the

33. *See* Edward J. Imwinkelried, *Should the Courts Incorporate a Best Evidence Rule into the Standard Determining the Admissibility of Scientific Testimony?: Enough Is Enough Even When It Is Not the Best*, 50 CASE W. RES. L. REV. 19, 33 (1999).

34. *See* David L. Faigman et al., *How Good Is Good Enough?: Expert Evidence Under* Daubert *and* Kumho, 50 CASE W. RES. L. REV. 645, 667 (2000); Imwinkelried, *supra* note 33, at 28.

35. Moreover, observational studies cannot necessarily fill the gap in the behavioral area, because of problems with determining the ground truth. Rape trauma syndrome is an example. In determining the effects of rape for purposes of distinguishing between rape victims and complainants making false claims of rape, the best comparison would be between women who accurately claim to have been raped and women who falsely claim to have been raped. This comparison is not feasible, so a comparison of the symptoms of women who report being raped and those who report not being raped is substituted for it.

36. Edward J. Imwinkelried, *Flawed Expert Testimony: Striking the Right Balance in Admissibility Standards*, 18 CRIM. JUST. 28, 29 (2003).

37. For studies on helpfulness of expert testimony about eyewitnessing conditions, see BRIAN L. CUTLER & STEVEN D. PENROD, MISTAKEN IDENTIFICATION: THE EYEWITNESS, PSYCHOLOGY AND THE LAW 19–54 (1995). *See also* Roger C. Park, *Eyewitness Identification: Expert Witnesses Are Not the Only Solution*, 2 LAW PROBABILITY & RISK 305, 306–07 (2003).

precedential value of decisions about whether the expertise meets standards of validity, make appellate review more difficult, and perhaps lead the judge into subjective assessments of lay testimony that are best left to the jury.

Another factor to be considered is whether the fact-finder will make a comparison of the trace and the source in the absence of expert testimony. For example, when the proffered testimony is bullet lead analysis, there will be no such comparison if the expertise is excluded. The information about the chemical composition of the bullets will simply not come in, and no one will make a comparison between the composition of the crime scene bullets and the composition of the bullets found on the defendant. On the other hand, if FDE testimony is excluded, the task of making the comparison between a questioned document and known exemplars can still be done. The jury will do it without the aid (or hindrance) of FDE testimony. Of course, it would be possible to prevent such comparisons altogether, and Jennifer Mnookin has called attention to early cases that did exclude comparison of handwriting evidence,[38] but this course seems unlikely in modern practice. Exclusion might have unforeseen side effects (the jury might wonder what was being concealed and hold it against one of the parties).[39] Moreover, the current Federal Rules of Evidence seem to contemplate that the trier of fact will compare handwriting, though it is not clear that they absolutely require it.[40]

In place of exclusion, one can attempt to reduce the prejudicial

38. Jennifer L. Mnookin, *Scripting Expertise: The History of Handwriting Identification Evidence and the Judicial Construction of Reliability*, 87 VA. L. REV. 1723, 1764–66 (2001) (citing nineteenth century cases excluding handwriting exemplars offered solely for the purpose of allowing the jury to make comparisons of handwriting).

39. *See* Old Chief v. United States, 519 U.S. 172, 188 (1997) (citing Stephen A. Saltzburg, *A Special Aspect of Relevance: Countering Negative Inferences Associated with the Absence of Evidence*, 66 CAL. L. REV. 1011, 1019 (1978) ("If [jurors'] expectations are not satisfied, triers of fact may penalize the party who disappoints them by drawing a negative inference against that party.")).

40. Federal Rule of Evidence 901(b)(3) provides that authentication may be accomplished by "[c]omparison by the trier of fact or by expert witnesses with specimens which have been authenticated." The Advisory Committee Note expresses disapproval of "common law restrictions upon the technique of proving or disproving the genuineness of a disputed specimen of handwriting through comparison with a genuine specimen, by either the testimony of expert witnesses or direct viewing by the triers themselves." FED. R. EVID. 901 advisory committee's note. The restrictions referred to were reservation to the judge of the question of genuineness of exemplars and "imposition of an unusually high standard of persuasion." *Id.* Of course, the advisory committee note was not enacted by Congress, and is merely a guide; moreover, its permissive attitude towards experts is arguably obsolete in light of the requirements of Federal Rule of Evidence 702, codifying the *Daubert* case. *See* United States v. Saelee, 162 F. Supp. 2d 1097, 1101 (D. Alaska 2001). However, it would seem odd to prevent comparisons by the trier of fact in view of this language and the provision in Federal Rule of Evidence 901(b)(2) allowing nonexperts to compare handwriting and give opinions. Federal Rule of Evidence 901(b)(2) specifically provides for authentication of handwriting by "[n]onexpert opinion as to the genuineness of handwriting, based upon familiarity not acquired for purposes of the litigation."

effect of exaggerated expertise with procedural devices. Examples include instructions warning about the flaws of the expertise, restrictions on expert use of scientific terms, and restrictions on testimony about certain inferences—for example, allowing testimony only about similarities between the trace and the exemplar, while prohibiting testimony about the ultimate conclusion that they came from the same source.[41] Procedural devices can also be used in an attempt to combat hidden bias, improve accuracy, or enhance the adversary system's ability to reveal the defects of experience-based testimony. Examples include requiring blind comparisons of the trace and exemplar,[42]evidence "line-ups,"[43] and case-specific proficiency tests.

II. THE PARTICULAR CASE OF SIGNATURE AUTHENTICATION

The seminal law review article on handwriting identification is *Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification "Expertise"*(*Exorcism* article).[44] In a well-known passage, the article states that "[a] rather generous reading of the data would be that in 45% of the reports forensic document examiners reached the correct finding, in 36% they erred partially or completely, and in 19% they were unable to draw a conclusion."[45] The authors add that after excluding a test that was unreasonably easy, a less generous conclusion would be that the examiners were correct 36% of the time, erred partially or completely 42% of the time, and were unable to reach a conclusion 22% of the time.[46]

The *Exorcism* article was a helpful wake-up call, one that provided plenty of ammunition for questioning claims of expertise that are either global (FDEs can do everything from detecting forgery to telling who did the forgery) or exaggerated (FDEs are error-free). Subsequent work by handwriting identification skeptics has taken into account additional data, including studies comparing expert-to-lay performance and comparing expert-to-chance performance.[47] It has also provided a more task-specific analysis of the expertise, distinguishing between different skills, such as the skills of signature authentication, of attributing authorship to disguised writing, and of attributing authorship to hand

41. *See* United States v. Hines, 55 F. Supp. 2d 62, 70–71 (D. Mass. 1999); United States v. Starzecpyzel, 880 F. Supp. 1027, 1030, 1051 (S.D.N.Y. 1995).

42. Risinger et al., *supra* note 20, at 45–47.

43. *Id.* at 47–50.

44. D. Michael Risinger et al., *Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification "Expertise,"* 137 U. PA. L. REV. 731 (1989).

45. *Id.* at 747.

46. *Id.* at 748.

47. *See* Risinger, *supra* note 5, at 379; D. Michael Risinger & Michael J. Saks, *Science and Nonscience in the Courts:* Daubert *Meets Handwriting Identification Expertise*, 82 IOWA L. REV. 21, 41–64 (1996).

printing.[48]

This Article will also attempt a task-specific analysis in conformity with the mandate of the *Kumho Tire* case.[49] The specific task that I will evaluate is that of determining whether a questioned signature is genuine. This task is treated as a distinct one by FDEs, which seems plausible in view of the differences between signatures and other writing, and between detecting simulation in signatures and the harder task of assigning authorship to disguised or simulated writing.[50] The skill could be broken down into sub-skills, for example, the subtasks of detecting forgery by skilled forgers versus naive forgers, or of detecting simulations of complex versus simple signatures.[51] However, too much subcategorization would impede the development of precedent and impose unrealistic demands on researchers.

By signature authentication, I am referring to the task of determining whether a questioned signature is the genuine, naturally-written signature of the person named in the signature.[52] A signature might be inauthentic in a number of ways. A person might write another's name as his "signature," without any attempt to imitate the other person's handwriting,[53] or the inauthentic signature might be an intentional imitation, either by tracing or by freehand imitation. Tracing could be done in several ways—for example, by placing the later-questioned document on top of the genuine signature with a light source behind the genuine-signature model, or by using the genuine-signature model to make indentations on a paper under the model and then inking the indentations. There are more exotic possibilities, such as machine-generated imitations and "guided hand" signatures produced by someone aiding or controlling the hand of the person named in the signature.

---

48. Risinger, *supra* note 5, at 379.

49. *See* Kumho Tire Co. v. Carmichael, 526 U.S. 137, 141 (1999); Risinger, *supra* note 5, at 389.

50. As Professor Risinger notes, standard FDE theory holds that it is much more difficult to attribute authorship to a forgery than it is to determine that a signature is not genuine. Risinger, *supra* note 5, at 394 n.13 (citing ALBERT S. OSBORN, QUESTIONED DOCUMENTS 286–87 (2d ed. 1929) [hereinafter OSBORN, 1929 edition]); *see also* WILSON R. HARRISON, SUSPECT DOCUMENTS: THEIR SCIENTIFIC EXAMINATION 374 (1966); Ordway Hilton, *Can the Forger Be Identified from His Handwriting?*, 43 J. CRIM. L. & CRIMINOLOGY 547, 547, 548, 555 (1953); sources cited *infra* note 215.

51. These are indeed just examples. Other conceivable subtasks would vary according to the number of exemplars available, whether the exemplars were demand exemplars (written with knowledge that they might be used for forensic identification) or course-of-business exemplars, or whether the person named in the questioned signature was in a different physical condition (older, sicker, in more pain) than when that person's exemplars were collected.

52. I say "named in the signature" despite the fact that some signatures are logos that cannot really be said to contain discernible letters of a name. The reports on the studies that I will describe give me no reason to think that logo signatures were predominant or even strongly represented in the studies.

53. This might occur because he lacks a model or because he does not think the signature will ever be questioned.

The first problem for a commentator is to hit upon a handy way of describing the different types of errors that can be made in signature authentication. The use of the familiar terms "false negatives" and "false positives," or "false hits" and "false alarms," would be technically accurate (if terms are defined) but possibly confusing. The reason is that some readers might intuitively regard a decision falsely saying that a signature was genuine as a "false positive," while others would think of a decision falsely saying that a signature was a forgery as a "false positive." In cases in which the evidence is being used in a criminal prosecution, we tend to think of a "false positive" as a decision that incorrectly identifies an innocent person as a criminal, and a "false negative" as one that incorrectly exonerates a guilty person. Using that convention, when the crime charged is forgery, a decision that mistakenly identified the signature as not being authentic would be a "false positive." But if another type of crime were charged (say, drug dealing, when the issue is whether a Western Union money order signed with the name of the defendant was signed by the defendant), then a decision that falsely identified the signature as being genuine would be a "false positive."

In an attempt at uniformity, I will use the following terminology to describe errors in signature authentication. The error of saying that a signature is genuine when it is not genuine will be described as "false authentication." The error of saying that a signature is not genuine when it is in fact genuine will be described as "false elimination." In making the latter error, the subject decided that the person named in the signature did not sign it (eliminated him), when in fact the person named in the signature did in fact sign it.[54]

First, I will examine two proficiency tests involving signature authentication administered by Collaborative Testing Services (CTS).[55]

---

54. One could argue for different terminology. For example, Moshe Kam uses the terms "false authentication" and "false simulation-detection" for the two types of error. Moshe Kam et al., *Signature Authentication by Forensic Document Examiners*, 46 J. FORENSIC SCIS. 884, 886 (2001). I have opted for the term "false elimination" for two reasons. First, it is shorter. Second, the studies that I will be describing (with a notable exception) do not specifically ask the subjects to distinguish between genuine signatures and "simulated" signatures. *See* Jodi Sita et al., *Forensic Handwriting Examiners' Expertise for Signature Comparison*, 47 J. FORENSIC SCIS. 1117, 1118 (2002). Instead, they ask whether the signature is genuine or not, or some very similar question. In addition, there is a difference between the concepts of "false elimination" and "false simulation-detection." A signature might be forged without simulation (the forger does not know what the genuine signature looks like, only the name of the person he is seeking to stand for) or simulated without forgery (seeking deniability, the person named in the signature makes an obvious tracing of his own signature, or creates a freehand disguise of his handwriting while writing his own name—for example, a right-handed person writes his name with his left hand, and tries to inject differences in letter slant and character formation).

55. *See* COLLABORATIVE TESTING SERVS., INC., CRIME LABORATORY PROFICIENCY TESTING PROGRAM: QUESTIONED DOCUMENTS ANALYSIS, REPORT NO. 88-5 (1988) [hereinafter CTS REPORT NO. 88-5] (on file with author); COLLABORATIVE TESTING SERVS., INC., CRIME LABORATORY PROFICIENCY TESTING PROGRAM: QUESTIONED DOCUMENTS ANALYSIS, REPORT NO. 85-8 (1985) [hereinafter CTS REPORT NO. 85-8] (on file with author). CTS is a private company that tests the performance of forensic and

These proficiency tests do not compare expert and lay performance, and there are various differences between the test environment and actual casework,[56] so they are of limited help in deciding whether experts will help jurors. Nonetheless, extreme results on proficiency testing could say something about the field. Were the experts to perform no better than chance, then their testimony should be excluded.[57] If intuition and introspection tell us that the task is easy and the experts surprise us with their poor performance, that is also a reason to doubt that their testimony would be helpful. Of course, experts might also surprise with a flawless performance on a task that looks hard, in which case we should be more receptive to what they have to say, though we would still have to bear in mind that we are both using intuition and ignoring differences between the test situation and actual casework. As in other decisions about the generalizability of test results, there is a good deal of room for subjective judgment.

A. THE 1985 CTS TEST

I will start with the 1985 CTS test.[58] It is worth some discussion because, depending on how one views the results, it suggests a high error rate. It is also one of the most famous proficiency tests, since it was discussed in the original *Exorcism* article,[59] in the Galbraiths' critique of that article,[60] and in the Risinger response to that critique.[61]

The test givers created twelve checks with signatures. (Apparently all of the signatures were written as "Elizabeth J. Drinkwalter," though the test materials are not totally explicit on this point.)[62] The same person

---

industrial labs. When the Forensic Science Foundation received grants from the U.S. Justice Department in the 1970s to test the proficiency of crime labs, CTS became a subcontractor and developed tests. *See* Thornton & Peterson, *supra* note 18, at 41 n.4. After the end of the federal grant, it continued on a fee-for-service basis. *Id.* The *Exorcism* article examined a number of additional proficiency tests in reaching the conclusions in the quoted paragraphs above. *See supra* notes 44–46 and accompanying text. While the *Exorcism* article was examining all tested aspects of handwriting expertise; I am limiting myself to the specific task of signature authentication.

56. *See infra* Part III.

57. For examples of difficult tasks on which experts performed no better than what would be expected by chance, see Oliver Galbraith III et al., *The Principle of the "Drunkard's Search" as a Proxy for Scientific Analysis: The Misuse of Handwriting Test Data in a Law Journal Article*, 1 INT'L J. FORENSIC DOCUMENT EXAMINERS 7, 14–16 (1995).

58. CTS REPORT No. 85-8, *supra* note 55. Professor Michael Risinger kindly provided me with copies of the CTS summary reports on this test and on the 1988 CTS test discussed below. The summary reports are the feedback given collectively to participating laboratories. They contain brief introductory remarks, the scenarios posed in the tests, the correct answers, and the results for individual laboratories (laboratory anonymity is protected by a coding system). Reports on the tests are not otherwise published. Risinger, *supra* note 5, at 521–22.

59. Risinger et al., *supra* note 44, at 745–76.

60. Galbraith III et al., *supra* note 57, at 7.

61. Risinger, *supra* note 5, at 383.

62. CTS REPORT No. 85-8, *supra* note 55, at 4–9. The proposition in text is this Author's inference from reading the narrative answers of the text takers.

signed the signature on checks #3 and #8.[63] All of the other checks were signed by different people. Checks #6 and #11 were (depending on the packet) either tracings or freehand simulations of the signatures on checks #3 and #8. All of the other checks were apparently written in the natural hand of the subjects, without any attempt to disguise the hand or imitate the hand of someone else.[64] The test takers were asked "[w]hich, if any, of the twelve checks were written by the same person?"[65] The correct answer was "3 and 8."[66]

Forty-one percent of the respondents answered that checks #3 and #8 (and only those checks) were written by the same person; 31% were not able to reach a conclusion; and the others reached a conclusion, but made one or more mistakes.[67]

These results can be characterized in different ways. One way is to say that the experts were correct only 41% of the time. A more generous way would be to discard the inconclusives, and say that the experts were correct 59% of the time.[68] (Many of the labs that gave inconclusive responses complained about the nature of the materials. The labs had to base their conclusions upon the twelve checks alone, and were not able to request additional exemplars of the signature of Elizabeth J. Drinkwalter.)[69]

Either way, the results indicate that the experts are not infallible. But, of course, expertise doesn't have to be infallible to be helpful. And the error rate can be viewed in many different ways, depending on how one slices the subtasks. An even more generous way of viewing the results would be to disaggregate the specific decisions made by the experts in each comparison of hands.

By analogy, suppose a teacher gives a multiple-choice test containing fifty questions. There are different ways that results could be reported. One could calculate the percentage of students who got any of the fifty questions wrong, and report that as the error rate. A more customary approach would be to treat each question as a separate task, and report the error rate as the mean percentage of questions answered incorrectly.

With that in mind, it is worth contemplating what a document examiner had to do in order to achieve a completely correct answer on

63. *Id.* at 2 tbl.1.

64. This is the Author's inference from the description of the task in the manufacturer's information. *See id.*

65. *Id.* at 2 tbl.2.

66. *Id.*

67. *Id.*

68. *See* Galbraith III et al., *supra* note 57, at 11, 14. The Galbraiths calculate the percentage correct in alternative ways, first by discarding inconclusives (their method 1, yielding 59% correct), then by analyzing the inconclusive results to see whether the document examiners gave qualified opinions of authorship (their method 2, yielding 55% correct). *Id.*

69. *See* CTS REPORT No. 85-8, *supra* note 55, at 1.

the 1985 CTS test. The document examiners received twelve signatures and were asked to identify the ones that had common authorship. The completely correct answer was that checks #3 and #8 were written by the same person, but none of the other checks had common authorship.[70] Making the test more difficult, check #6 was a freehand simulation of the signature on one of the matching checks (#3 or #8), and check #11 was a tracing one of the matching check signatures.[71]

Based only on signatures (no extra exemplars), the document examiner would have to decide that two (and only two) of the signatures were produced by the same person, just on the basis of the two signatures themselves without the aid of additional writing samples.[72] Then the examiner would have to reject all the other signatures as being nonmatches, including two signatures that were freehand forgeries or tracings of the two signatures that were written by the same person. Since the examiners were free to say that all or none of the signatures matched each other,[73] there were 4,084 possible answers to the question, and the random chance of getting a completely correct answer would be 1/4084.[74] The experts clearly outperformed chance on this task. To put it another way: the experts who gave a completely correct answer performed with 100% accuracy on many subtasks. Even those who gave partly wrong answers performed correctly on most of the subtasks. If each of the subtasks had been considered a separate decision, then the mean accuracy rate across decisions would have been much higher than that derived from counting all the subtasks as if they were one decision.

To do this, let us first place the errors into two categories: (1) Making a mistaken same-source decision when the signatures come from different sources (the signatures were written by different people, but the FDE decides that they were written by the same person); and (2) Making a mistaken different-source identification when the signatures come from the same source (the signatures were written by the same person, but the FDE decides they were written by different individuals).

We can now disaggregate the tasks by conceiving of the 1985 test as one that requires the FDEs to make sixty-six decisions about whether pairs of signatures were written by the same person. Thus, the FDE subjects had to decide whether signature 1 had the same source as signature 2, whether signature 1 had the same source as signature 3, etc. For the 1985 test, the completely correct answer was that one of the signature pairs had the same source (signatures 3 and 8 were written by the same person), and that none of the other signature pairs had the

70. *See supra* note 66.

71. *See* CTS REPORT No. 85-8, *supra* note 55, at 2 tbl.1.

72. *See id.* at 1.

73. *See id.*

74. Galbraith III et al., *supra* note 57, at 14.

same source.

Viewed this way, the results of the FDE decisions as to whether pairs of signatures came from the same source were as follows:[75]

TABLE I: 1985 CTS TEST

GROUND TRUTH	DECISION	
	SAME SOURCE	NOT SAME SOURCE
SAME SOURCE	21	1
NOT SAME SOURCE	52	1,378

Thus, when the ground truth was that two signatures came from the same source, the FDEs made twenty-one correct decisions and one erroneous decision. When the ground truth was that the two signatures did not come from the same source, the FDEs made 1,378 correct decisions and fifty-two erroneous decisions (treating eight "possibly the same source" decisions as incorrect).

Determining whether the signatures come from the same source is the same task as determining whether a signature is authentic (though with only one exemplar, a handicap for the FDEs). For example, when an FDE compares signature 3 to signature 8 and decides they were written by the same person, that is the same thing as treating signature 3 as a known exemplar of the person's signature and saying that signature 8 is genuine. Therefore, the two errors can fairly be described in terms of the terminology used throughout this Article: the "false authentication error" and the "false elimination error."

Using this terminology, the experts had the following error rates:

(1) False authentication error rate: 3.8% (of 1,430 pairs that did not come from the same source, the FDEs erroneously decided that they came from the same source fifty two times, and correctly decided that they came from different sources 1,378 times).

(2) False elimination error rate: 4.5% (of twenty two pairs that did come from the same source, the FDEs erroneously decided that they came from different sources one time, and correctly decided that they were from the same source twenty one times).[76]

---

75. *See* CTS REPORT No. 85-8, *supra* note 55, at 2–3 tbl.2. I have excluded responses that declined to express an opinion (e.g., "inconclusive" answers). This eliminated the data from ten responding labs, leaving twenty-two respondents. That decision increased both the error rate and the correct answer rate (compared with the alternatives of giving percentages for three categories: "correct," "inconclusive," and "error"). Where an answer said certain pairs of signatures came from the same source and certain other pairs "possibly" came from the same source, I have treated the "possibly" answers as being authentications. That was a decision that disfavored the document examiners. Had the "possibly" answers been treated as exclusions, there would have been eight fewer same-source errors.

76. *See id.*

Some of the decisions were arguably easy, because they involved finding that the natural handwriting of different people was different (though it must be remembered that the FDEs had no exemplars, and looked only at the twelve "signatures" written by eleven different individuals). We can try to isolate the more difficult task (determining whether a simulated signature was authentic) by breaking the false authentication errors into two categories:

(1) False authentication of a nongenuine signature made with no attempt at simulation.[77]

(2) False authentication of a nongenuine signature made with an attempt at simulation.[78]

Returning to the figures in Table 1, when the ground truth was that the signatures came from different sources, there were fifty-two false authentication errors and 1,378 correct exclusions. But let us now focus only on the pairs that came from different sources and in which the signature writer made an attempt at simulation by tracing or free-hand simulation.[79] Each FDE faced five such pairs, so the twenty-two respondents together faced a total of 110 such pairs. In classifying these 110 intentional simulations, the FDE respondents made eighteen errors (sixteen errors if two "possible" answers are excluded).[80] Thus, on this task the FDEs produced ninety-two correct answers and eighteen errors, for a 16% error rate.

Is this error rate too high? Answering that question requires a healthy dollop of intuition. To me, the error rate seems acceptable, even admirable, when we consider what the FDEs were being asked to do. They were being asked to decide, on the basis of two "signatures" containing the words "Elizabeth J. Drinkwalter," whether another "signature" in the name of "Elizabeth J. Drinkwalter" was written by the same author, when the third "signature" was in fact a conscious imitation.[81] The FDEs had to detect, on the basis of this scanty

77. Technically, this should be described as "false authentication of handwriting of someone signing another's name as the same handwriting as that contained on specimens provided by another person who was signing a name not her own," an arguably harder task. *See infra* note 161.

78. Technically, this should be described as "false authentication of the handwriting of someone simulating specimens signed by a person who was not signing her own name," an arguably harder task. *See infra* note 161.

79. Each of the twenty-two responding FDEs faced five such pairs. CTS REPORT No. 85-8, *supra* note 55, at 2–3 tbl.2. Depending on the test package, #6 was a simulation of either #3 or #8, and #11 was a simulation of either #3 or #8. *Id.* at 2–3 tbls.1 & 2. Since #3 and #8 were written by the same person, I am treating a simulation of #3 as also being a simulation of #8, and vice versa. Also, simulations of a signature from the same source are treated as simulations of each other (if you disagree with this judgment, count four pairs instead of five). Thus, the five simulation pairs were 3-6, 3-11, 6-8, 6-11, and 8-11.

80. *See id.* at 2–3 tbl.2.

81. *See id.* at 2–3 tbls.1 & 2. In order to have two signatures as known exemplars, the FDEs needed first to determine that Drinkwalter Signature #3 was written by the same person as

information, whether the third signature was different because of intrawriter variation or because it was written by a different person. The fact that they were fooled by the imitation only 16% of the time seems rather good, considering the paucity of information.[82]

Three obstacles made the tasks harder for the FDEs. One was the absence of exemplars. Another was that the test takers worked from photographs rather than originals. A third was that although all of the "signatures" were specimens containing the name "Elizabeth J. Drinkwalter," none of the signers, including the person who signed #3 and #8, was named "Elizabeth J. Drinkwalter."[83] If writers perform more awkwardly (or with less individuality) when signing an unfamiliar name than when signing a signature developed over a lifetime, then the use of a false name may have been misleading. (In fact, one of the responses stated: "Checks 3 and 8 could have been written by the same writer not signing one's own name," though the expert thought that it was more likely that they were written by different writers having similar styles.)[84] In its report on the test, the testing agency conceded that "[i]n hindsight, it may have been better if the name used had been the true name of the person who wrote #3 and #8."[85]

In view of these difficulties, the test manufacturer's assessment is one reasonable way of viewing the results. After noting the problem that the test takers were unable to request additional signatures, the CTS report states:

> Despite these difficulties, thirteen laboratories gave responses consistent with the manufacturer's information and an additional two were only marginally different. Ten laboratories produced inconclusive responses, largely because of the nature of the samples. Seven labs gave responses at least partially inconsistent with the manufacturer's information. Most of these results were a result of failure to recognize the simulation in #6 and/or the tracing in #11. The [Proficiency Advisory Committee] was pleased to see that the majority of labs did pick up on at least one of these simulations.[86]

While the 1985 CTS test does not establish the value of the expertise, it does not rebut it, either. The experts did much better than chance, and even those who were partly wrong might have been helpful to juries by pointing out features of the signatures. To borrow a term from some of the test takers, the results were inconclusive.

Drinkwalter signature no. 8.

82. In fact, considering the lack of exemplars, one can make a case for treating "inconclusive" answers as also being correct, in which case the error rate would shrink to 11%.

83. *Id.* at 1.

84. *Id.* at 9 (answer of respondent 829).

85. *Id.* at 1.

86. *Id.*

B. THE 1988 CTS TEST

The Forensic Science Foundation administered another proficiency test relevant to signature authentication skills in 1988.[87] The 1988 CTS test was based on a fact scenario in which complaints were received from physicians' offices about missing shipments of narcotics.[88] The delivery service produced four receipts containing four purported signatures in the name of the secretary for each physician.[89] The respective secretaries denied that they wrote the signatures.[90] Demand exemplars from the four secretaries and the driver of the delivery truck were taken.[91] The delivery service also provided two other receipts signed by unknown persons.[92]

The report on the 1988 CTS test does not state whether any of the nongenuine signatures were simulations using as a model the genuine signature. It is possible that the "delivery driver" Richard D. Osbourn did not know what the signatures of the secretaries looked like, or at least did not trace them or make free-hand simulations.

Like the report on the 1985 CTS test, the report on the 1988 test does not provide the stimulus materials (the photos of the signatures and exemplars). I have inferred which names were used for the signatures on the receipts from the comments of the test takers. [93]

A fuller description of the 1988 CTS test appears in Appendix 3. Here, I will consider only the portions of the test that bear on the task that is the subject of this Article: signature authentication.

On that particular task, a complete set of correct answers would have been:

(1) The receipt Q1, signed with the name "Sharon D. Clayborne," was not signed by Sharon D. Clayborne.

(2) The receipt Q2, signed with the name "Lisa D. Bridgeforth," was indeed signed by Lisa D. Bridgeforth.

(3) The receipt Q3, signed with the name "Cynthia Y. Boone," was not signed by Cynthia Y. Boone.

(4) The receipt Q4, signed with the name "Joanna Neuman," was indeed signed by Joanna Neuman.

The text of the CTS report does not itself state the percentage of correct answers on each of the four tasks listed above. However, it does state the answers given by each lab. I have hand tabulated those answers,

87. CTS REPORT No. 88-5, *supra* note 55.

88. *Id.* at 40.

89. *Id.*

90. *Id.*

91. *Id.*

92. *Id.*

93. I checked my work against Professor Risinger's tabulation of names. Our results were the same. *See* Risinger, *supra* note 5, at 522.

with the results set forth below. Where the correct answer was that the receipt was indeed signed by the person whose name appears in it, I have counted "was written by" and "was probably written by" as being correct answers. Where the correct answer was that the receipt was not signed by the person whose name appears in it, I have counted "was not written by" and "was probably not written by" as being correct answers. Where the test taker left the answer space blank but indicated elsewhere that the questioned signature was or probably was written by someone other than the person whose name appears in the signature, I counted that answer as a determination that the signature was not authentic.

Placing the four signature authentication tasks into the two categories used in this Article, and counting only "called answers,"[94] the error rate was:

(1) False authentication error: 1% (1/96)

(2) False elimination error: 6% (4/71)

The test takers had various complaints about the test.[95] For example, they were furnished with photographs of the signatures instead of the originals, and some of the takers complained that this made analysis of certain features more difficult.[96] Others noted that the exemplars were made upon request (i.e., they were demand exemplars, made at one sitting by a subject who would realize that they were going to be used for handwriting comparison) and wished they also had course-of-business exemplars (examples of natural handwriting made over a period of time without forensic tests in mind).[97]

After the end of the Justice Department Program for proficiency testing, the CTS administered a number of other proficiency tests as a private company on a fee-for-service basis.[98] If inconclusives are discarded, these tests indicate a correct answer rate exceeding 90% on signature authentication tasks, and on some of the comparisons 100% of the FDEs were correct.[99] The exception is one of the tasks on the 2001

94. For a more detailed tabulation, see Appendix 3. "Called" answers are ones that state whether the signature is authentic or not, as opposed to "inconclusive" answers and other answers that do not decide the authentication issue. In calculating the percentage of correct answers, my inclination is to exclude the "inconclusive" and "other" answers. With one exception, the "other" answers either leaned toward the right answer or called for more evidence. The "inconclusive" answers may have also been based on the view that the evidence was insufficient. That is not necessarily a wrong answer. For example, if Ms. Neuman did in fact have a simple signature that she signed with significant variation, it seems reasonable to call for more exemplars or to conclude that the evidence was not enough for a conclusion.

95. CTS REPORT No. 88-5, *supra* note 55, at 34–39 tbl.5.

96. *Id.* at 34–38 tbl.5.

97. *See id.*

98. *See supra* note 55.

99. *See* Risinger, *supra* note 5, at 558–65 (discussin COLLABORATIVE TESTING SERVS., QUESTIONED DOCUMENTS ANALYSIS, REPORT NO. 92-6 (1992); COLLABORATIVE TESTING SERVS., QUESTIONED DOCUMENTS EXAMINATION, REPORT NO. 9406 (1994); COLLABORATIVE TESTING SERVS., QUESTIONED

CTS proficiency test.[100] There, the three questioned documents were entries on a "sign-in log" that included a signature purporting to be the signature of Kenny Bania as well as handwritten numbers denoting date and time.[101] The FDEs who took the test had been provided with exemplars from Bania and from two other persons, but not from the person who simulated Bania's signature.[102]

Entries Q1 and Q3 had in fact been written by Kenny Bania.[103] Entry Q2 was a freehand simulation of Kenny Bania's signature and number writing.[104] The FDEs were 100% correct (with no inconclusive answers) in saying that Kenny Bania had in fact written Q1 and Q3.[105] However, their results were far less good when they reached the signature authentication task required in Q2.[106] On that task, seventy-five respondents (78.5%) correctly answered that Bania did not or probably did not write the questioned signature,[107] but twenty (21.5%) identified

---

DOCUMENTS EXAMINATION, REPORT NO. 9606 (1996); COLLABORATIVE TESTING SERVS., FORENSIC TESTING PROGRAM, HANDWRITING EXAMINATION, REPORT NO. 9714 (1997); COLLABORATIVE TESTING SERVS., FORENSIC TESTING PROGRAM: HANDWRITING EXAMINATION, REPORT NO. 9814 (1998); COLLABORATIVE TESTING SERVS., FORENSIC TESTING PROGRAM: HANDWRITING EXAMINATION, TEST NO. 99-524 (1999); COLLABORATIVE TESTING SERVS., FORENSIC TESTING PROGRAM: HANDWRITING EXAMINATION, TEST NO. 00-524 (2000); COLLABORATIVE TESTING SERVS., FORENSIC TESTING PROGRAM: HANDWRITING EXAMINATION, TEST NO. 01-524 (2001) [hereinafter CTS TEST NO. 01-524] (on file with author); COLLABORATIVE TESTING SERVS., FORENSIC TESTING PROGRAM: HANDWRITING EXAMINATION, TEST NO. 02-524 (2003)).

100. *See* CTS TEST NO. 01-524, *supra* note 99. In my description, I have extracted the parts of the task that relate to the narrow task of signature authentication as I have defined it (determining whether the person named in the signature wrote the signature), and not to the other task presented in the test—that of deciding whether a simulated signature should be attributed to someone else. The attribution task may have made the authentication task more difficult on Q2, because none of the exemplars of other persons apparently suspected of simulating the signature were exemplars of the person who did simulate the signature, so the examinee who assumed that either the real Kenny Bania or one of the other suspects was the one who wrote the signature would have been led astray. This problem could, of course, also cause error in ordinary casework.

101. *Id.* at 2–4.

102. *Id.* at 2.

103. *Id.*

104. *Id.*

105. On Q1, 94% (123/131) accurately responded that Kenny Bania was the writer, and 6% (8/131) responded that he probably wrote the entry. *Id.* at 7. On Q3, 93% (122/131) accurately responded that Kenny Bania was the writer, and 7% (9/131) responded that he probably wrote the entry. *Id.* at 13.

106. Twenty FDEs positively eliminated him, fifty-five opted for probable elimination. *Id.* at 10. One reason why it may have been harder for the FDEs to eliminate than it was for them to identify is the possibility that the person named in the signature actually wrote the signature, but in a disguised hand so as to be able to later deny writing it. *Id.* at 3. Thus the detection of signs of simulation does not conclusively exclude the possibility that the person whose name appears in the signature did in fact write the signature. Sometimes the other case facts might make writing in a disguised hand highly unlikely, and this may be a situation in which it would be appropriate, at some point, to give the FDE access to other case facts even if those facts are initially screened so as to prevent observer bias. For example, cases involving allegedly forged wills are probably ones in which it seems highly unlikely that the real decedent will have signed her own signature in a disguised hand.

107. I am describing it as a questioned signature even though the signature was actually accompanied by a few numbers. To be precise, the task was whether to authenticate the signature and its accompanying numbers, but the signature seems to have provided most of the relevant information

him as the author of the questioned signature and there were thirty-six inconclusive answers.[108]

Thus, the error rates for the called opinions were:

(1) False authentication error: 22% (20/95)

(2) False elimination error: 0% (0/262)

Not surprisingly, this proficiency test shows a much higher rate of false authentication error than the tests in which the signers apparently made no attempt to imitate a genuine signature. One lesson that can be drawn from the proficiency tests is that a forger who imitates a genuine signature will sometimes succeed.

The 2001 Australian study by Found and Rogers also provides error rate data for intentional simulations.[109] There, on the task of detecting genuine signatures as genuine, the FDEs were 92% correct, 6% inconclusive, and 2% affirmatively wrong.[110] On the task of detecting that nongenuine signatures were simulations, they were 43% correct, 53% inconclusive, and 4% affirmatively wrong.[111] (The nongenuine signatures were freehand forgeries by naive forgers who were allowed to practice using a genuine signature as a model.)[112] Discarding inconclusives, the accuracy rate was:[113]

(1) Simulated as simulated: 91.5 %

(2) Genuine as genuine: 98.2%

In other words:

(1) False authentication error: 8.5%

(2) False elimination error: 1.8%

A third category of signature in the 2001 Found & Rogers study was the "disguised" signature.[114] A disguised signature is a signature in the true name of the specimen writer, but written in a disguised hand.[115] It is the kind of signature that might be written, for example, by a receipt signer who was signing her true name, but who wanted to make the signature look as if someone else had signed it, so that she could later

---

for performing the task.

108. *See* CTS TEST NO. 01-524, *supra* note 99, at 10.

109. My description of the Found and Rogers study comes from examination of a printout of a presentation setting forth their results. *See* BRYAN FOUND & DOUG ROGERS, REVISION AND CORRECTIVE ACTION PACKAGE: SIGNATURE TRIAL 2001 (on file with author). This study was distributed on CD-ROM by the Forensic Expertise Profiling Laboratory, School of Human Biosciences, La Trobe University, Australia, and provided to this Author by Michael Risinger.

110. *Id.* at 26 tbl.2. The raw scores were 1,628 correct, 30 wrong, and 105 inconclusive. *Id.* at 25 tbl.1.

111. *Id.* at 26 tbl.2. The raw scores were 2,840 correct, 265 wrong, and 3455 inconclusive.

112. *Id.* at 12.

113. *Id.* at 26 tbl.2.

114. *Id.* at 8.

115. *Id.* at 9.

deny having signed it. Here the results were much more mixed. Of the FDEs tested, 29.6% correctly assigned authorship to the specimen writer, 23.9% incorrectly said she was not (or probably not) the author, and 46.4% answered "inconclusive."[116] Discarding "inconclusives" and using only called opinions, 55.3% were correct and 44.7% incorrect.[117] This data suggests that assigning authorship to disguised handwriting is a much harder task than determining whether a signature was written in the natural and usual style of the signer. In cases where self-disguise is a realistic possibility, perhaps signature authentication experts should only be allowed to state an opinion about whether (1) the signature is the naturally written signature of the person named in the signature; or (2) there are signs of disguise or simulation. This restriction would mean that the experts would not be allowed to assign authorship to disguised or simulated writing.

Without seeing the raw stimulus materials (the questioned signatures and known exemplars), it is difficult make even an intuitive judgment about how hard the tests I have discussed were. They might be too hard, too easy, or just right. (Even if I had access to the original stimulus materials, my intuitive judgments about difficulty would probably be wrong; all FDE tasks seem hard to me.) Moreover, the proficiency tests discussed above do not compare expert performance with lay performance, and hence do not tell us much about whether it would be better to use an expert or have the jury do the signature authentication alone. Third, participation in proficiency testing is voluntary, and it is difficult to estimate the effect of selection bias. The labs that requested and returned materials might have been the best labs, a reasonable inference in view of the fact that they were interested in testing their performance. But they might have been the least busy labs, or labs that wanted to test trainees, or labs that had something to prove and want to be able to say that they had passed proficiency tests. The labs that requested but did not return materials might have been the ones that found the test too hard, or they might have intended from the beginning to use the tests for future training when they were not so busy. Finally, there are problems with generalizing the test results to actual casework. Perhaps the test takers were more careful in doing the tests than they would be in actual casework, or more prone to give "inconclusive" opinions than in actual casework. In actual casework, where the FDEs typically know the other evidence in the case and the result that the prosecution would like to have, biasing effects might well convert some of the "inconclusives" to positive answers (or even change the positive answer to their opposites).[118] On the other hand, in actual

116. *Id.* at 26 tbl.2.

117. *Id.* at 26 tbl.2.

118. See Risinger et al., *supra* note 20, at 21.

casework the FDEs might benefit from being able to request additional exemplars or from being able to use original signatures and exemplars instead of photographs.

Another reason to hesitate before using existing proficiency test data inferences about the validity of the expertise is the possible effect on the tests themselves. The current commercial test makers (CTS) eschew any claim that the test results are valuable in assessing the field, and in fact warn against using them for that purpose.[119] As noted above, one of the reasons why it is difficult to draw conclusions from proficiency test results is that the tests may be too hard or too easy. That quality is manipulable. A test maker could intentionally devise a test that would result in a 1% error rate or a 99% error rate. The test maker who is trying to actually test proficiency aims at something in between. The CTS test makers are aware of this factor (though their recent tests seem to have been on the easy end of the spectrum). This can be seen in their "summary comments" to Test No. 01-524 (2001), where they proudly quote the comment of a test taker who said, "This is an excellent problem. A mistake awaits anyone who is not cautious and thorough."[120] After noting that the test taker who made that flattering comment had a perfect score on the test, they note that another test taker who criticized the test as "too easy" was wrong on a crucial answer.[121]

That is exactly the attitude that we want the test makers to have. They should strive to create a reasonably difficult test that will cause mistakes by those who are not careful, thorough, and able. But if the tests are used by doubters to attack the field, the test takers are unlikely to compliment the tests for being difficult. They are more likely to demand easy tests.

A program of proficiency testing could be devised that would be both free of that danger and more informative in estimating an error rate. Ideally, the program would be administered by persons with a scientific background who are neutral in the sense that they have no personal interest in either defending or disparaging FDE expertise. Moreover, the program should be blind, meaning that the proficiency tests are presented as if they were ordinary casework, so that the test

---

119. *See, e.g.*, CTS REPORT No. 88-5, *supra* note 55, at intro.

> Since it is the laboratory's option how the samples are to be used (e.g., training exercise, known or blind proficiency testing, research and development of new techniques), the results compiled in the summary report are not intended to be an overview of the quality of work performed in the profession , and cannot be interpreted as such. . . . They are included for the benefit of participating laboratory directors to assist them with maintaining or enhancing the quality of results from their individual laboratories. These comments are not intended to reflect the general state of the art within the profession.

Id.

120. *See* CTS TEST No. 01-524, *supra* note 99, at 3.

121. *Id.*

takers do not know they are being tested.[122] However, blind testing that is extensive enough to justify statistically sound conclusions about error rates is likely to be expensive and difficult to implement.[123] One problem is making the testing actually blind, when labs are accustomed to interacting with investigators, for example by contributing investigative leads and calling for more evidence to help them perform their FDE comparisons.[124]

Even with an extensive blind proficiency program, there would still be unanswered questions. First, one would still not know how expert performance compares with lay performance. Second, the aggregate data, while useful, would not be definitive with regard to any particular lab—the lab might be better (or worse) than the typical lab, or it might be using newer methods (or discarding old ones). But perfection is unattainable, and these issues could be explored in testimony. If enough proficiency testing were done on enough specific subtasks, the results would be useful, either in helping judges make the *Daubert* decision or in informing juries of the possibilities of expert error.

III. Studies Comparing Expert and Lay Performance

A. Description of the Kam and Sita Studies

I will now discuss the studies comparing expert and lay performance on the task of signature authentication.[125] Probably the best-known study, at least in the United States, is the study by Moshe Kam and colleagues, *Signature Authentication by Forensic Document Examiners*, published in 2001.[126] In May 1998, Dr. Kam and his colleagues conducted a test of signature authentication with sixty-nine FDE subjects and fifty lay subjects.[127] These subjects examined signatures that had been generated by other participants.[128] The stimuli-generating participants were graduate and undergraduate students at Drexel University who were

122. *See* Joseph L. Peterson et al., *The Feasibility of External Blind DNA Proficiency Testing: Background and Findings*, 48 J. Forensic Scis. 21, 26 (2003) (presenting data indicating that in a variety of forensic fields, blind testing yields fewer positive calls than open testing).

123. For similar problems with blind proficiency testing by DNA labs, see Margaret A. Berger, *Laboratory Error Seen Through the Lens of Science and Policy*, 30 U.C. Davis L. Rev. 1081, 1088–89 (1997).

124. *See* Risinger et al., *supra* note 20.

125. I have not included the pilot study comparing expert and lay performance set forth in Galbraith III et al., *supra* note 57, at 7, on grounds that it did not involve a signature authentication task.

126. Kam et al., *supra* note 54. Dr. Kam has done other studies comparing expert and lay performance on assignment of authorship to naturally written nonsignature handwriting, but those studies are not directly relevant to signature authentication, and they had methodological problems involving differential lay and expert incentives that were not present in the same degree in the 2001 study. For a description of these earlier studies, see Risinger, *supra* note 5, at 527–49.

127. Kam et al., *supra* note 54, at 884.

128. *Id.* at 885–86.

paid $25 for three hours' work.[129] Each of those participants provided twelve "freely and naturally executed" examples of his or her own signature.[130] (These were divided randomly into six-signature subsets.)[131] Seven other participants were hired to simulate those signatures "using the manual techniques described in a text of forensic document examination."[132] They were provided with tracing paper, carbon paper, flashlights, and overhead projectors.[133] The simulators had no known prior experience in simulating signatures.[134] Each simulator was given six genuine signatures and allowed as much time as they wanted to practice the simulation.[135] The investigators obtained professional FDEs from three different FDE conferences to act as FDE subjects.[136] The lay subjects were staff, faculty, and students from Drexel University.[137] All test takers were given a known set of six genuine signatures and told that they had been provided voluntarily by the signer in a single sitting.[138] They were also given an unknown set of six signatures (i.e., questioned signatures).[139] The unknown set could be all genuine signatures, all simulated signatures, or anything in between.[140]

Subjects were given decision choices of "identification," "strong probability [did write]," "elimination," and "strong probability did not write."[141] They were also given an inconclusive option.[142]

The FDE and lay performance compared as follows:[143]

TABLE II: KAM ET AL., 2001

GROUND TRUTH	DECISION					
	GENUINE		INCONCLUSIVE		NOT GENUINE	
	FDE	LAY	FDE	LAY	FDE	LAY
GENUINE	85.89%	70.00%	7.05%	4.30%	7.05%	26.10%
NOT GENUINE	0.49%	6.47%	3.45%	1.40%	96.06%	92.00%

129. *Id.* at 885.
130. *Id.*
131. *Id.*
132. *Id.* at 885 (citing W. HARRISON, FORGERY DETECTION: A PRACTICAL GUIDE (1964)).
133. *Id.* at 885.
134. *Id.*
135. *Id.*
136. *Id.* at 885–86.
137. *Id.* at 886.
138. *Id.*
139. *Id.*
140. *Id.*
141. *Id.*
142. *Id.*
143. *Id.* at 887. The differences in error rates for the FDE and lay subjects were statistically significant. *Id.*

TABLE III: KAM ET AL., 2001: CALLED DECISIONS

GROUND TRUTH	DECISION			
	GENUINE		NOT GENUINE	
	FDE	LAY	FDE	LAY
GENUINE	92.41%	72.84%	7.59%	27.16%
NOT GENUINE	0.51%	6.57%	99.49%	93.43%

These tables show that when the questioned signature was in fact genuine, the FDEs correctly said it was genuine 85.89% of the time, said "inconclusive" 7.05% of the time, and made the false elimination error in 7.05% of their decisions. If we exclude the inconclusives, then the FDEs were correct in 92.41% of decisions, and made the false elimination error in 7.59% of decisions. Lay subjects were correct on 72.84% of these calls, and made the false elimination error on 27.16% of them. When the questioned signature was in fact a simulation, the FDEs made the false authentication error 0.49% of the time, said "inconclusive" 3.45% of the time, and correctly designated it as nongenuine in 96.06% of the decisions. If we exclude the inconclusives, then the FDEs were correct in 99.49% of decisions, and incorrect in 0.51% of decisions. The comparable figures for lay subjects are 93.43% and 6.57%.

A 1999 Australia-New Zealand pilot study by Bryan Found, Jodi Sita, & Doug Rogers reached similar results, though with higher rates of inconclusive answers.[144] There, seven document examiners and eight lay persons were asked to make judgments about the authenticity of each of 150 questioned signatures.[145] The overall error rate of the lay subjects was 28%, compared to 2% for the document examiners.[146] The lay error rate for the false authentication error was 7%, compared to 0% for the document examiners.[147] The lay error rate for the false elimination error was 21%, as opposed to 2% for the document examiner group.[148] The document examiner group was considerably more conservative in making "calls" and had a much higher "inconclusive" rate than the lay group.[149] Although I have examined the published study, I cannot report exact figures for "inconclusive" answers because they are presented in a bar-graph table that does not make very fine distinctions. However, it appears that the FDEs gave about seventy "inconclusive" answers, compared to about forty "inconclusive" answers by the lay subjects. The differences between the lay errors and the FDE errors were statistically

144. Bryan Found et al., *The Development of a Program for Characterizing Forensic Handwriting Examiners' Expertise: Signature Examination Pilot Study*, 12 J. FORENSIC DOCUMENT EXAMINATION 69, 75–76 (1999).

145. *Id.* at 72.

146. *Id.* at 76.

147. *Id.*

148. *Id.*

149. *Id.*

significant, as were the differences between lay "inconclusives" and FDE "inconclusives."[150] The lay subjects actually gave more correct answers than the FDE subjects (about seventy eight as compared to about eighty two) but the difference was not statistically significant.[151]

The same authors followed up with a larger study in 2002.[152] (My comments from now on will concern the larger study, not the pilot study.) There, seventeen FDEs from five Australian and New Zealand government forensic laboratories and thirteen lay subjects participated.[153] Ten volunteers each executed thirty free-hand, natural signatures over a twelve-month period.[154] Twenty-five other volunteers made freehand simulations of those signatures, with as much time as they wanted to practice.[155] The simulators then submitted a "one-off" simulation executed on a specially marked sheet reserved for a single try, and a "best try" simulation that they thought to be their best forgery.[156] Test packages were prepared that called for the FDEs and the lay subjects to perform the signature authentication task.[157] A judgment of genuine, simulated, or inconclusive was elicited for each questioned signature.[158]

For reasons that are unclear, the authors of the Sita study did not compare lay performance to expert performance in the same categories used by Kam. The errors of false identification and false elimination were combined, as were the correct answers of detecting genuine as genuine and detecting simulated as simulated.[159] The results were as follows:

(1) FDE: 54.8% correct, 41.8% inconclusive, 3.4% wrong

(2) Lay: 57.1% correct, 23.6% inconclusive, 19.3% wrong

Note that the lay persons again actually made a higher percentage of correct decisions than the experts, though the results were not statistically significant. The experts were superior in avoiding error, to a statistically significant degree.[160]

If one counts only "called" opinions (that is, if inconclusives are excluded) then the error rates compare as follows:

(1) FDE: 5.8% errors

(2) Lay: 25.3% errors

150. *Id.* at 71.

151. The lay subjects had a higher percentage of both correct answers and wrong answers because the experts made more frequent use of the option of answering "inconclusive."

152. Sita et al., *supra* note 54, at 1117.

153. *Id.* at 1118.

154. *Id.*

155. *Id.*

156. *Id.*

157. *Id.*

158. *Id.*

159. *Id.* at 1119.

160. *Id.*

Differences between lay and expert error rates were statistically significant.

As in the Kam study, the FDE subjects in the Sita study made significantly more errors in false elimination (calling a genuine signature to be not genuine) than they did in false authentication (calling a simulated signature genuine).[161] The called error rate for false elimination was 12.2%, as opposed to 2.1% for false authentication.[162] The data comparing false elimination and false authentication for lay persons are not set forth in the report of the study.

The Kam study and the Sita study differ greatly in the percentage of "inconclusive" answers. It is not clear why this should be so, at least for the FDE subjects. (The lay subjects in the Sita study were given cautionary instructions that may have encouraged inconclusive answers.) There is no obvious difference in the test materials that can be detected by examining the investigators' published accounts, which do not include photographs of the questioned signatures and specimens. Kam collected a set of twelve freely and naturally executed genuine exemplars for each signature; Sita collected a set of fifteen genuine exemplars (presumably also freely and naturally executed) for each signature.[163] The simulators were naive simulators in both studies.[164] Perhaps the Australia/New Zealander document examiners in the Sita study required a higher degree of certainty before rendering an opinion than the American FDE examiners in the Kam study. Perhaps the Sita subjects, in "psyching out" the Sita test, thought that it was going to be a particularly difficult one. The Sita test materials hinted that some of the signatures might be disguised signatures by the person whose name was signed, and that one of the tasks involved might be to distinguish between a "natural" signature by the person whose name is written as the questioned signature and a "simulated" signature authored by the person whose name is written as the questioned signature.[165] A previous study by two of

161. *Id.* at 1118–19. For consistency of terminology, I am continuing to use the categories of "false authentication" and "false elimination," even though technically the errors might be described as "false missing of simulation" and "false detection of simulation." The reason is that the study asked the subjects to distinguish between "genuine" signatures and "simulated" signatures. They were instructed to answer "genuine" when "[t]he questioned signature is in your opinion written by the same person who wrote the standard signature group" and to answer "simulated" when "[t]he questioned signature is inconsistent with the standard signature group and displays features that you consider to be indicative of a copying process. Note that this term does not imply that the standard signature writer did not write it." *Id.* at 1118. This instruction seems hard to follow, even illogical (a person who traced his own signature would fall in both categories), and I am assuming that subjects called signatures "genuine" when they thought the signature had been written by the person named in the signature, and "simulated" when they thought the signature was not genuine.

162. *Id.* at 1120 tbl.5. The comparable called error rates in the 2001 Kam study were 7.59% for false elimination and 0.50% for false authentication. *See* Kam et al., *supra* note 54, at 887.

163. Sita et al., *supra* note 54, at 1118.

164. *Id.*

165. *Id.*

the same authors had included a task of detecting disguised signatures written by the specimen writer, a task that had a high error rate.[166] Or it may simply be that, for some reason not apparent on the face of the published studies, the task in the Sita study was harder than the task in the Kam study.

B. COMMENTS ON THE KAM AND SITA STUDIES

There are various ways to describe the difference between FDE and lay performance. The difference seems most impressive if one compares the ratio of errors. Table 4 shows the ratio for called opinions on the 2001 Kam study:

TABLE IV: KAM ET AL., 2001: ERROR RATIOS

ERROR	FDE	LAY	ERROR RATIO (LAY TO FDE)
FALSE AUTHENTICATION	0.51%	6.57%	12.9 TO 1
FALSE ELIMINATION	7.50%	27.16%	3.6 TO 1

The most dramatic ratio difference is in the false authentication error. The lay subjects were over thirteen times as likely to make the error of saying that a simulated document was genuine. But in absolute terms the error rate was not very high for either FDEs or lay subjects.

The Sita study, in its comparison between FDE and lay performance, does not distinguish between the two types of error.[167] The aggregate data are set forth in Table 5:

TABLE V: SITA ET AL., 2001: ERROR RATIOS

ERROR	FDE	LAY	ERROR RATIO (LAY TO FDE)
FALSE AUTHENTICATION OR FALSE ELIMINATION	3.4%	19.3%	5.7 TO 1

Both studies show FDEs making significantly fewer errors than the lay subjects. But the path from these studies to the conclusion that expert testimony would be helpful at trial is strewn with pitfalls.

First, there is the bothersome question of what to do with the "inconclusive" answers. The experts were far more likely to say "inconclusive" than the lay subjects in both studies. If "inconclusive" is considered a wrong answer, then the difference between expert and lay performance shrinks.

166. FOUND & ROGERS, *supra* note 109, at 8.

167. Sita et al., *supra* note 54, at 1120–21.

The question whether to consider "inconclusive" as a wrong answer is difficult, a difficulty that is compounded by the fact that an examination of the question would involve using the expertise to judge the expertise. A comparison with fingerprint identification may illustrate the point.

Suppose that a fingerprint expert is asked to make a comparison of a complete set of ten carefully rolled prints with another complete set of ten carefully rolled prints from the same subject. If the expert states that the comparison is "inconclusive," then it seems fair to deem that result erroneous (unless one is a strong skeptic of the value of any fingerprint comparisons). But suppose that the expert is asked to compare a sparse latent print with the subject's rolled print. Even if the latent print comes from the same subject, it seems perverse always to label an "inconclusive" answer as wrong. The "inconclusive" answer simply means that there is not enough data to make a decision. The expert may be correct in saying that there is not enough data even if the prints come from the same person. In fact, the more skeptical one is about the technique of identification, the less ready one should be to call an "inconclusive" answer wrong. Judging whether an inconclusive answer is "wrong" involves using expertise to determine whether the expert's answer was not up to standard.

In handwriting identification, one reason for an "inconclusive" answer may be that the expert has simply not been provided with enough exemplars. This was the reason given by FDEs for some of the "inconclusives" in the 1985 CTS proficiency test.[168] This might explain some of the "inconclusives" in the Kam 2001 and Sita 2002 studies, even though the FDEs were given six and fifteen exemplars in the respective studies. But there are other justifiable reasons for an inconclusive judgment. One is that the subject's signature has so much natural intrawriter variation that it is unusually difficult to make a judgment. Another is that the subject's signature is so simple and commonplace that there is too great a danger of interwriter similarity.[169] An expert who reaches the "inconclusive" conclusion on those grounds might deserve respect rather than criticism.[170] In part the answer depends upon that ever-present but elusive decision about who gets the benefit of the doubt. My intuition favors discarding the "inconclusive" answers, treating them

168. CTS REPORT No. 85-8, *supra* note 55, at 34.

169. *See* Sita et al., *supra* note 54, at 1122.

170. On the other hand, unless the test makers were intentionally asking FDEs whether signatures were genuine or not without providing enough information to make the judgment, the experts that the test makers consulted in devising the test must have thought that there was enough data to justify an answer. Of course, it is possible that the test makers simply devised a plan for making the stimulus materials (have people sign their signatures and other people imitate them) without giving any subsequent thought to whether the results provided enough information for a judgment of genuineness.

as neither right nor wrong answers, but as nonresponses. Under this approach, one counts only the "called" answers, the instances in which the FDEs or lay subjects came to a conclusion. This has the effect of increasing both the error rate and the correct answer rate (compared to the alternative of using three categories, "correct," "inconclusive," and "erroneous").

The document examiners may have had greater motivation to give inconclusive answers in the study environment than did the lay subjects. Discarding the "inconclusives" partly compensates for these dangers by increasing the error rate of subjects who use the "inconclusive" category (their error rate is a percentage of their total affirmative decisions, not a percentage of all three categories). Nonetheless, using the "inconclusive" category might still be a good strategic move for test takers when they are really in doubt (but leaning in one direction) and when the aimed-for error rate is very low. Perhaps the solution is to design tests in which lay people give as many "inconclusive" answers as FDEs. This could be done by fiddling with financial incentives, or by doing a separate measure in which the lay subjects rate their degree of certainty and the investigator uses the certainty measure as a way of assigning the desired percentage of lay answers to the inconclusive category.

The greater use of "inconclusives" by FDEs would not be a problem if one could be sure that the study conditions generalize to trial conditions. Recognition that comparisons are inconclusive would be a contribution that experts could make to the process, counteracting a lay tendency to jump to conclusions. The problem is that FDEs may be more cautious in test conditions than they are in real casework, especially if in real casework they are exposed to extraneous information. This is part of the larger problem of external validity discussed below in Part III.D.

C. Challenges to Internal Validity

There is a danger, in both the Kam 2001 study and the Sita 2002 study, that something other than expertise caused the differences in performance. Possibilities include differences in age or background that are not related to the expertise. The most disturbing possibility, however, is that differences in motivation caused the difference in performance, by causing the FDEs to take the task more seriously than the lay subjects and to work longer and harder at it.

First, the FDEs may have feared that word of mistakes in individual performance might leak out, leading either to impeachment at trial or to their competency being called into question by supervisors or peers. Both studies appear to have made an effort to prevent this result. Second, FDEs are aware of the challenges that have been made to the field on which they depend for their livelihood, and they would want the study outcome to show them performing well. Moreover, their self-

esteem would take a blow if the test indicated that they were not skilled at what they do for a living, whereas the lay subjects are not likely to care very much whether they are good at signature identification tasks.

Dr. Kam made a commendable effort to motivate the lay subjects with money.[171] The lay subjects received $8 for each correct decision and lost $8 for each affirmative error (indecision had a consequence of $0, $4, or –$4 depending upon the incentive group to which the lay persons belonged; the researchers did not detect any difference in performance due to this difference).[172] While this addresses the problem, it probably falls short of the incentives for document examiners. The Sita 2002 study did not even try to solve the incentive problem. It gave no performance-based monetary rewards or punishments to subjects.[173]

The incentive problem is hard to solve, short of giving huge monetary rewards or using complicated deceptions (such as using a signature authentication task as a mock entrance exam for an FDE training program). The problem is, however, mitigated to some degree in the legal policy context. If FDEs are not allowed to testify about signature authentication tasks, then fact finders (judges or jurors) whose careers are not at stake will make the signature comparisons. They may not be as highly motivated as FDEs taking proficiency tests either; their motivation may more closely resemble that of the lay subjects in the Kam tests. Of course, jurors want to do justice, and they might be more motivated than a student, even a scholarship student, who is susceptible to a swing of $16 if she decides to make an affirmative determination of authenticity. But jurors want to go home. Self-interest is usually thought to be a pretty good motivator. The Kam subjects had that spur and jurors do not (or if they do, it points in the direction of going home). As is so often the case, applying experimental results to a real-life legal context depends ultimately upon unsystematic inductions from experiences and introspection—what some would call "intuition."

Another challenge is self-selection among document examiners. It may be that only the best examiners participate in studies such as those conducted by Kam and Sita. Of course, there are other possibilities, and there is a good dose of speculation in assessing this danger.

D. Challenges to External Validity

Even if the difference in results between the FDEs and the lay subjects in the Kam and Sita studies is due to expertise and not to some other factor, the results may not generalize to the pertinent legal context.

One reason is that FDEs may perform differently on proficiency

171. Kam et al., *supra* note 54, at 886.

172. *Id.*

173. Sita et al., *supra* note 54, at 1123.

tests than in actual casework. They may be more highly motivated to be careful and thorough in proficiency tests. Moreover, in proficiency tests their judgments are not affected by extraneous information. In actual casework, FDEs are likely to have access to other information. For example, they might know that a witness will testify that the defendant in a forgery case made self-incriminating statements, or that other evidence links the signer of a Western Union receipt to a drug ring. Their desire to be a member of the prosecution team may also impair their judgment in actual casework. Among other things, they may be less conservative in giving "inconclusives" when the answer toward which they lean will help the prosecution team.

These concerns are serious, though irreducibly speculative. It seems likely that document examiners take proficiency tests more seriously than real casework, but there is no conclusive proof. Some of them may have a sense of justice that makes them try harder in real cases. Some of them may be taking the proficiency exams under circumstances in which there is no career harm if they make mistakes. And some real cases may have higher consequences, because of the danger that document examiners will be proven wrong. Where the document examiner gives an opinion during the investigation process, for example, subsequent information developed by the police may prove the opinion wrong. And there is always a possibility that after trial, information will turn up that disproves their position, as appears to have happened in the Dreyfus case.[174]

Finally, when generalizing from the results of the Kam study to actual trial situations, one must ask whether the relatively accurate performance of lay subjects on the task of correctly identifying nongenuine signatures means that expert testimony is not worth a candle at trial when the FDE is proffered for the purpose of opining that a signature is not genuine. If the ground truth is that the signature is not genuine, and the 2001 Kam test is used as the gauge of accuracy, then the expert will contribute the correct decision 96% of the time, the lay person 92%. If the prosecution is correct, then the lay jury will reach the correct decision 92% of the time.[175] The difference between 92% and 96% may not be enough to justify the cost and time of involving an expert, especially when one considers that the trial situation may be a better one for lay decision making than the experimental situation. In the trial situation, jurors have the benefit of group deliberation and arguments of counsel. The arguments of counsel could be informed by consultation with FDEs even if the FDE testimony were excluded. Points

174. *See* D.H. Kaye, *Revising* Dreyfus: *A More Complete Account of a Trial by Mathematics*, 91 MINN. L. REV. 825, 827–28.

175. *See supra* tbl.2. If inconclusives are discarded, the lay subjects were correct on 93.4% of their calls, and the FDEs were correct on 99.5% of their calls. *See supra* tbl.3.

that were overlooked in the solitary comparisons done by lay subjects in the Kam and Sita studies might thus come to the jurors' attention.

Possibly the simulated signatures used in the 2001 Kam study were crude and easy to detect, creating a ceiling effect that did not leave the experts much room to excel. However, although error rates were low for both FDEs and lay subjects, the lay subjects did make thirteen times as many affirmative errors as the FDEs.[176] At any rate, if the FDEs are in fact correct 96% of the time (or 99.5% if only called opinions are counted)[177] then it is hard to say that their testimony would be very prejudicial. The only question would be whether the testimony would be worth the cost. In view of the fact that the alternative methods of proof also have drawbacks and that the parties may incur costs by consulting with FDEs even if the testimony is not admissible,[178] this does not seem to be a compelling reason for excluding FDE testimony. It may, however, be an additional reason for restricting FDE testimony to pointing out similarities and differences, instead of allowing FDEs to state an ultimate conclusion.[179]

E. The 1975 Conrad Study

A third article comparing expert and lay performance, Wolfgang Conrad's 1975 study of German document examiners,[180] provides less reason for being optimistic about the abilities of FDEs on the signature authentication task. Conrad compared the results of twenty five "publicly contracted and sworn experts" ranging in age from twenty-four to eighty-two years with that of lay persons with no handwriting background and of university students who had taken handwriting courses.[181] The participants were presented with materials containing genuine and forged signatures and asked to reach a judgment about whether the signatures were genuine. Some of the forgeries had been produced by freehand procedure and others by tracing. Conrad reported that "the quality of the forgeries on the whole exceeds the level usually found in

176. *See supra* tbl.4. The error rate for FDEs was 0.49%, for lay subjects 6.47%. The reason why the correct answers and errors do not sum to 100% is that subjects were also allowed to answer "inconclusive." On this task, 3.45% of the FDE answers were inconclusive, compared to 1.40% of the lay answers. *See supra* tbl.2.

177. *See infra* note 182.

178. *See* discussion *infra* Part IV.

179. *See* discussion *infra* Part IV.

180. Wolfgang Conrad, *Empirische Untersuchungen über die Urteilsgüte verschiedener Gruppen von Laien und Sachverständigen bei der Unterscheidung authentischer und gefälschter Unterschriften* , 156 Archiv Für Kriminologie 169–83 (1975), *translated in* Empirical Studies Regarding the Quality of Assessments of Various Groups of Lay Persons and Experts in Differentiating Between Authentic and Forged Signatures (Peter Bernhardt trans.) (unpublished manuscript on file with author). I am grateful to Michael Saks for providing me with the English translation by Peter Bernhardt of Dr. Conrad's article.

181. *Id.* at 2 (English translation).

the judicial practice."[182]

The most startling finding was that the university students had a lower aggregate error rate than the generally-qualified experts (though not lower than experts who were specially selected for accuracy on a prior performance test). It is, however, unclear what weight should be given to this result. First, the University students were students whose qualifications were "successful completion of the classes Handwriting Psychology I and II as well as Handwriting Comparison I and II."[183] Students with those qualifications may be more analogous to well-educated but inexperienced FDEs with systematic training in handwriting comparison than to American jurors. Second, there were only six subjects in the student comparison group, and the difference between their performance and that of the experts was not statistically significant.

The twenty-five generally-qualified experts[184] substantially outperformed one hundred "name owners fairly representing the adult population of the Federal Republic of Germany"[185] and one hundred lay persons who apparently were also representative of adult West Germans.[186] On the only comparison of the generally-qualified experts with these lay groups, the twenty-five experts had a 14.7% error rate compared to 34.4% for the general-population lay persons and 25% for the actual "name owners" themselves.[187] However, when the generally-qualified FDEs were compared to lay persons with similar occupations and education,[188] the difference in performance was minor: the general group of experts had a 14.7% error rate, and the similar-education-and-profession group of lay participants had an error rate of 20%.[189] When only called opinions were used, the difference in error rates between these two groups increased, but not very much. The general group of experts had an error rate of 10%, compared to an error rate of 18% for the similar lay subjects.[190] (A small group of specially screened experts

182. *Id.* at 3.

183. *Id.* at 4 tbl.1.

184. A small group of six specially screened experts did substantially better than the generally-qualified experts, and a group of sixteen specially qualified experts did somewhat better. *Id.* at 4 tbl.1.

185. *Id.* at 2.

186. *Id.* at 6.

187. *Id.* at 6 tbl.1. Inconclusives were counted as errors on this comparison, and the experts may have suffered from use of this decision criteria, if, like the FDEs in the Kam and Sita studies, they were more prone to give inconclusive answers than were lay participants.

188. The educational level of the twenty-five experts was as follows: three had finished elementary school, six had received their general certificate of education from secondary school, and sixteen had passed their "Abitur," or graduation exam from Gymnasium qualifying them for university entrance. Of the "Abitur" experts, eleven had completed university studies. *Id.* at 2.

189. *Id.* at 6 tbl.1 (counting inconclusives as wrong answers).

190. *See id.* at 6 tbl.1. I am using Conrad's "error grade II" criteria, under which opinions were called with a safety grade of probable or higher. In other words, I have excluded inconclusives but

had an error rate of 2.8%, and a group of specially qualified experts had an error rate of 9.4%).[191]

In the terminology used in this Article, comparing the called opinions of the generally-qualified group of experts with the lay participants most similar to them:[192]

(1) False authentication error: FDEs 12% Lay 12%

(2) False elimination error: FDEs 8% Lay 24%

This is hardly encouraging news for FDEs, particularly the false authentication error rate: the generally-qualified group of FDEs were fooled by the well-executed forgeries used in the Conrad study to the same extent as were the similar lay participants. The lay false elimination error rate was, however, three times greater than the FDE rate. (The small group of six specially screened experts had no errors on either task, but for all we know the same statistic might be derived if specially talented lay people had been tested.)

The Conrad study is evidence against FDE expertise, but it is not clear how much weight should be given to it. Perhaps there are differences in signature characteristics or FDE characteristics that make it difficult to generalize from a 1975 German study to other times or other countries. Moreover, the false authentication error statistics set forth above were based on a comparison of the decisions of twenty-five lay persons who examined three forged signatures with the decisions of twenty-five FDEs who examined three forged signatures. If the number of subjects or signatures had been greater, differences between lay and expert performance might have been detected.[193]

IV. PROCEDURAL SAFEGUARDS

The issues in the legal context are further complicated by the fact that admission and exclusion of expert opinions are not the only choices. The expert opinion can be admitted with procedural safeguards that are designed to protect against prejudice. The trial judge can impose these conditions in a hearing on a motion in limine in which one of the parties seeks to exclude the testimony.

In the *Hines* case, for example, the expert was allowed to testify about similarities and differences, but not to reach an ultimate conclusion about authorship.[194] In the context of signature authentication,

---

accepted called opinions with any degree of certainty.

191. *Id.*

192. *Id.* at 8–9 tbls.2 & 3 (error grade II comparison).

193. Each group had a total of nine errors. Outlier performance by one or two members of either group could have had a material impact on the percentage performance figures.

194. United States. v. Hines, 55 F. Supp. 2d 62, 70–71 (D. Mass. 1999). I was not a case involving signature authentication, but rather a case in which the expert would have attributed authorship of a robbery note used in a bank robbery to the defendant. *Id.* at 62. Signatures are regarded as special by

this could translate into allowing the expert to point out similarities and differences between the questioned signature and the exemplars. The signs of simulation would be among the differences. For example, if the questioned signature had blunt endings, extraneous patching, tremor, and a "drawn" appearance, and the exemplars did not, then the expert could point out those differences.[195]

Of course, the jury is almost certain to infer the expert's view from the exegesis on similarities and differences. This is especially so if the expert is allowed, for example, to state that blunt endings are a sign of forgery and to explain why (the alternative here would be to have the witness point out the blunt endings and have counsel make the argument). Nonetheless, the similarities and differences approach may be useful in helping the jury to reason. It prohibits a specific statement of individualization—that this person signed the document to the exclusion of all other possible signers, based on the expert's many years of experience. It lays open the nature of the reasoning process involved—because of many similarities, it seems likely that the signature is genuine, or because of many differences, it seems likely it is not.

One can think of other possible restrictions along the same lines. A drastic restriction would be to prohibit the expert from making any statements about a handwriting characteristic being rare or common—

---

the document examiner community, and at any rate there is likely to be a greater danger of disguise when composing a note that one knows that police investigators will scrutinize (such as a "stick-up" note). Assigning authorship to disguised handwriting is a difficult task.

195. These are among the signs of simulation mentioned in the FDE literature. *See* ORDWAY HILTON, SCIENTIFIC EXAMINATION OF QUESTIONED DOCUMENTS (rev. ed. 1982); OSBORN, 1929 edition, *supra* note 50; SCIENTIFIC EXAMINATION OF QUESTIONED DOCUMENTS 94 (Jan Seaman Kelly & Brian S. Lindbloom eds., 2d ed. 2006). Kelly and Lindbloom, sum up by saying that "[h]esitation, unnatural pen lifts, patching, tremor, uncertainty of movement as portrayed by abrupt changes in the direction of the line, and a stilted, drawn quality devoid of free, normal writing movements combine to reveal the defective nature of a poor-quality simulation." *Id.* Their more complete catalogue of suspicious signs also includes: pooling of ink, *id.* at 82; uniformly heavy stroke, *id.* at 89; presence of a lightly drawn outline, *id.* at 90; and smearing caused by erasure of an outline, *id.* at 91. When one of the exemplars known to be genuine is suspected as being the model for a tracing, then wandering away from the stroke of the genuine signature and returning to the common track is indicative of tracing, as is indentation in the genuine signature thought to have been used as a model. *Id.* at 91–92. Suspicious signs noted in OSBORN, 1910 edition, *supra* note 28, include: evidence of erasure, *id.* at 45; pencil outline or impression on paper under the signature, *id.* at 77; uneven, slow, wiggly, rough, haltering, clumsy, hesitating, irregular writing in questioned document when authentic signatures were smooth and natural, *id.* at 109–11; blunt as opposed to tapering terminal strokes, *id.* at 114–15; deviation from uniform strokes (tremors), *id.* at 117; interruptions of movement in direct curves or straight lines, *id.* at 117; deviation from pen pressure or alignment of genuine exemplars, *id.* at 123–25, 133; and hesitation, abnormal changes of direction, inconsistent pen pressure, unnatural movement interruption, pen lifts, retouching, *id.* at 267–69. Where a genuine signature may have been used as a model, exact replication of the genuine signature in the questioned one is a sign of tracing. *Id.* at 276–77. See also OSBORN, 1929 edition, *supra* note 50, where signs of forgery include: tremor not attributed to age, weakness, or illiteracy; *id.* at 110; disconnections or pen-lifts, *id.* at 114; alignment not consistent with rest of document, *id.* at 115; and patching, *id.* at 129. For an empirical study of the frequency of some of these features in forged signatures, *see infra* note 209.

that is, no statements other than that supported by systematic empirical research. This seems a rather extreme approach, restricting even descriptive, summarizational testimony.

Another approach, adopted in the *Starzecpyzel* case, is to prohibit the expert from assuming the mantle of science by using scientific language.[196] Thus, one might prevent the expert from calling himself a scientist, from referring to his "laboratory," or from testifying to misleadingly precise degrees of certainty.[197]

The judge can also attempt to protect against prejudice by instructing the jury that FDE expertise is not scientific. For example, in the *Starzecpyzel* case, the judge prepared instructions that said that forensic document examination was a practical skill, not a scientific one, and that "despite anything you may hear or have heard, it does not have the demonstrable certainty that some sciences have."[198] While not prohibiting the use of all scientific terminology, the court also sought to guard against its impressiveness by instructing that "although forensic document examiners may work in 'laboratories,' and may rely on textbooks with titles like 'The Scientific Examination of Documents,' forensic document examiners are not scientists-they are more like artisans, that is, skilled craftsmen."[199]

Other safeguards could be directed at making the trial experience more similar to the studies comparing expert and lay performance. For example, in the studies the experts are blind to extraneous case information and they have no reason to expect that a certain result will aid a party with whom they are aligned.[200] The most practical way to achieve a similar result in casework would be through reforming the procedures of crime labs.[201] However, it is possible that an inventive judge could do something in a particular case. For example, in a case meriting the time and trouble, the judge could require that the questioned signature and exemplars be submitted to an expert not previously involved in the case, for example at the FBI crime lab, with a case file that is devoid of extraneous case information that would affect the expert's opinion.

Another procedural safeguard would be the admission of "anti-expert" expert testimony about the shortcomings and limits of FDE expertise. One can hope that efforts to educate judges about the deficiencies of forensic science will at least lead to a more liberal attitude

196. United States v. Starzecpyzel, 880 F. Supp. 1027, 1029, 1036 (S.D.N.Y. 1995).

197. *Id.* at 1048. The *Starzecpyzel* court prohibited the expert from using a misleadingly precise nine-point scale to express degrees of certainty.

198. *Id.* at 1051.

199. *Id.* at 1050.

200. *See* studies cited *supra* notes 126, 144, 180 (Kam, Sita, and Conrad studies).

201. *See* Risinger et al., *supra* note 20, at 10.

toward experts who would point out those flaws to the jury. Familiarity with the scientific method, not training in the practice of the craft, should be the criteria by which the qualifications of an "anti-expert" are judged.

Judges should also take care to make sure that the party opposing the FDE testimony has had adequate discovery. That is another way in which FDE testimony might affirmatively contribute to justice, as opposed to leaving the comparison to argument of counsel. If an FDE makes a determination ahead of time and produces supporting reasons, then that allows more thoughtful scrutiny and a greater chance for adversarial testing. Unfair surprise would be more likely if the points were left to argument of counsel. A predicate of this benefit, however, is full and complete discovery. If the FDE determines that the signature is not genuine, for example, the FDE should describe in detail the signs of simulation and the differences between the exemplars and the questioned signature.

Another tool would be to give the opposing party the option of foregoing cross-examination of the expert, and instead require the proponent to substitute the written report turned over on discovery for the oral testimony of the expert. This procedure would ensure a complete written report and would also ameliorate the danger that extraneous personal characteristics of the expert would have too much influence on the jury. The party harmed by the report has a right to confront the witness who wrote it, but this right can be waived. Adoption of this approach would, however, require judges to rethink the traditional position about the value of demeanor evidence to the jury.[202]

Finally, there is the question whether case-specific proficiency testing ought to be allowed. This is a difficult question, for two reasons. First, it might not provide enough protection for the party opposing the expert—let us say, the defense in a criminal case. The reason is that the defense counsel may be reluctant, except in cases of real desperation, to take the chance. Giving the expert a chance to pass a proficiency test devised by the defense bears a family resemblance to asking O.J. Simpson to try on the bloody glove. Secondly, defense-initiated proficiency testing may just inject confusion and collateral issues. It is possible to concoct a proficiency test that is unfairly hard, by finding multiple signatures from different sources that are unusually simple and similar, or by hiring an experienced forger to make imitations. Moreover, case-specific proficiency testing would add cost and money, since it would not be fair to have the expert make decisions on the spot, without the kind of care that the expert would spend preparing the opinion in the

202. For skeptical assessments of demeanor evidence, see Roger C. Park, *Empirical Evaluation of the Hearsay Rule*, *in* ESSAYS FOR COLIN TAPPER 91 (Peter Mirfield & Roger Smith eds., 2003); ALDERT VRII, DETECTING LIES AND DECEIT (2000); and Olin Guy Wellborn III, *Demeanor*, 76 Cornell L. Rev. 1075 (1991).

case. Most likely there would need to be some kind of judicial supervision, perhaps even a court-appointed proficiency tester, and this would add a layer of complexity and supervision that is probably unrealistic, especially since it would require fairly early judicial involvement in a case that might, after all, never go to trial.

CONCLUSION

It is time to reach a conclusion about what judges should do with FDE testimony in signature authentication cases. Before doing so, however, I should emphasize that my conclusion applies only to that type of case. I agree with Professor Risinger[203] that, in the spirit of *Kumho Tire*, one cannot make a global judgment about the admissibility of FDE testimony.[204] It depends on the task at hand. The task at hand that I have discussed is signature authentication.[205] There are other tasks that are much harder and that have less warrant in the empirical literature. These include tasks such as assigning authorship to hand printing or assigning authorship to disguised handwriting. It is one thing to say that a signature is forged; quite another to say who the forger is.[206]

Even within the territory of signature authentication, there are, of course, different levels of difficulty. The task is made more difficult if there are few exemplars, or if the signature is a simple one with little artistry, or if there is a possibility of a skilled forger. But here I think we must compromise a bit, if only in pity for the poor judge who must make a task-specific determination. Too much is sacrificed if we try to subdivide the task further. Endless subdivision would destroy the precedential value of decisions about the admissibility of forensic science evidence.

---

203. Risinger, *supra* note 5, at 449.

204. *Kumho Tire* calls for a task-specific judgment. *See* Kumho Tire Co. v. Carmichael, 526 U.S. 137, 153–54 (1999).

> [T]he specific issue before the court was not the reasonableness in general of a tire expert's use of a visual and tactile inspection to determine whether overdeflection had caused the tire's tread to separate from its steel-belted carcass. Rather, it was the reasonableness of using such an approach, along with Carlson's particular method of analyzing the data thereby obtained, to draw a conclusion regarding the particular matter to which the expert testimony was directly relevant. That matter concerned the likelihood that a defect in the tire at issue caused its tread to separate from its carcass.

Id.

205. Leading FDE authors have identified signature authentication as a special subtask. *See* HILTON, *supra* note 195, at 173 ("[T]he identification of signatures constitutes a specialized branch of handwriting examination."); JOE NICKELL, DETECTING FORGERY: FORENSIC INVESTIGATION OF DOCUMENTS 59 (1996); OSBORN, 1929 edition, *supra* note 50, at 18–94, 384; *see also* sources cited *supra* note 49.

206. *See* OSBORN, 1929 edition, *supra* note 50. In fact, cases in which it is possible that the questioned signature was written by the person named in the signature, but in a disguised hand (so that the person could later deny signing the document) also present this different, more difficult task—that of assigning true authorship to a disguised hand. That is another reason to limit testimony to similarities and differences and to signs of simulation.

When deciding whether signature authentication expertise should be admitted, the alternative should always be in mind. The choice is between two imperfect processes. Either the jury will compare hands after the expert has done so, or the jury will do so on its own. Either process has plenty of opportunity for error.

Visualize the trial in which no expert testimony is admitted, but the question whether a signature is genuine is crucial. The exemplars and the questioned signature would be sponsored by lay witnesses who have personal knowledge. The sponsoring witness might be an investigating officer, or it might simply be whoever witnessed the questioned signature and the exemplars. The proponent does not always get to choose its witnesses.

If the case is important enough, it is probable that the proponent will have consulted an FDE, even if the FDE's testimony is likely to be excluded. The FDE will have examined the questioned signature and the exemplars in her place of work, using magnification where needed. The FDE will likely have prepared blow-ups with annotations showing signs of forgery. If the allegation is that the questioned signature is a tracing of one of the exemplars, then the FDE will have prepared exhibits that facilitate comparison of the questioned signature and the exemplar thought to have served as a model, for example by superimposing one upon the other on a document camera that projects in the courtroom.

If allowed to testify, the FDE might point out differences in handwriting style—letter formation, slant, adherence to a baseline, etc.,—between the questioned signature and the exemplars. She would also probably identify suspicious features of the questioned signature. The features thought to be suspicious include tremor, pen lifts, blunt endings, abrupt changes in direction, pooling of ink, minuscule patching, and a "stilted, drawn quality devoid of free, normal writing movements."[207] Too much similarity to an exemplar is also suspicious, as indicative of tracing.[208] Examples of trial testimony about traced and freehand simulations appear in Appendices 1 and 2. A well-read FDE might also reference a degree of systematic empirical support for the hypotheses that touchups, blunt endings, tremor splices, pen lifts, and superimposability are suspicious features.[209]

---

207. HILTON, *supra* note 195, at 185.

208. *See id.*; *see also* S.C. Leung et al., *Forgery II-Tracing*, 38 J. FORENSIC SCIS. 413, 423–24 (1993) (reporting the results of an empirical study, using a writing pressure meter, of signatures traced by 189 subjects). All the signatures traced were "highlighted by the pressure of a slow measured stoke accompanied with hesitation, pen pause and absence of vigor and spontaneity"; the investigators measured "superimposability" (overlapping strokes of traced simulations with the model used) and found that "the probability that a questioned signature has been produced by tracing from another (genuine) signature is related to the superimposability of the two." *Id.*

209. See Black et al., *supra* note 14, and authorities cited therein. Black et al. collected 177 genuine signatures and 620 simulated signatures, and asked an FDE to count pen lifts, blunt endings, tremor,

Now, suppose that all expert testimony is excluded. The jury will do the comparison. If the comparison is important, it would not make sense to simply introduce the questioned signature and the exemplars, without comment. Jurors might miss important points, such as minuscule patching. They might give too much importance to differences that could be explained (the tremor was caused by illness, not tracing) or rebutted with additional evidence (genuine handwriting not introduced by the proponent has a feature different from the proponent's exemplars and similar to the questioned signature). At any rate, the adversaries are unlikely to leave comparisons to the jury, nor should they be required to do so. The specific facts and the inferences to be drawn from them should be stated in open court, where the parties have a chance to contest them with arguments and counter evidence.

What about just having the lawyers make the comparisons and argue about handwriting the same way that they argue about other circumstantial evidence? In their closing arguments, lawyers could certainly make many of the points that an FDE would make. They are legitimate inferences from circumstantial evidence. For example, it would be reasonable for a lawyer to argue that the genuine signatures were smooth and flowing, whereas the questioned signature, though it had the same letter formations, was suspicious because its wavering line quality was a sign of slow writing, which was a sign of forgery. (In Osborn's words: "A straight line is not only the shortest distance between two points but also the quickest distance.")[210] But this method could lead to surprise and an underdeveloped treatment of the subject. It raises dangers of giving too much importance to similarities and differences that could be explained or rebutted with other evidence. FDE testimony might be less dangerous, especially if it were conditioned upon a detailed expert report and rigorous pretrial discovery.

Another way to try to avoid these dangers would be to have a "show and tell" lay witness who, during the trial, points out similarities, differences and suspicious signs. (Sometimes the witness who sponsors the exhibits will be able to do this, but it would be unfair always to saddle the party with that witness, who might be inarticulate or uncooperative.) One possible complication is that, if a lay witness were to make the sort of minute comparisons described in Appendix 2, the jury might think the

---

splices, and touch ups. *Id.* at 20. These signs of simulation occurred much more frequently in the simulations than in the genuine signatures. *Id.* at 21. The comparative percentages of suspicious features in the genuine to simulated signatures were: 18% to 52% (pen lifts), 17% to 73% (blunt endings), 10% to 62% (tremor), 0% to 99% (splices), 0% to 19% (touch ups). *Id.* at 21. The value of the study is somewhat diminished by the fact that all of the genuine signatures were produced by one person (the simulations were produced freehand by thirty-one different participants). *Id.* at 19. On superimposability of tracings, see Leung et al., *supra* note 208, at 408–10.

210. OSBORN, 1929 edition, *supra* note 50, at 107. For other references to tremor as a suspicious sign, see *id.* at 110; HILTON, *supra* note 195, at 185; and NICKELL, *supra* note 205, at 68.

testimony strange or contrived, and wonder whether something was being hidden. Who is this person? If he's not an expert, how did he notice all those details? Aren't there experts in this field? Why don't they have an expert testify? To prevent this sort of reasoning, it might be necessary for the court to explain that expert testimony is not being permitted because the experts have not been proven to be better than jurors, or at least that expert testimony is not being permitted because the judge is not satisfied that it would be sufficiently helpful.

There would still be the question of how to handle the avowedly nonexpert "show and tell" witness. Would the nonexpert be allowed to reveal that he had consulted with an expert? Allowing this might frustrate the goal of preventing a jury from relying upon untrustworthy expertise, and allow the expert's implied opinion to have some effect even though the expert is not presented for cross-examination. Yet failing to say this might make the nonexpert look peculiar, if he notes many things and uses equipment such as microscopes.

Another cost of limiting testimony to lay witness and leaving inferences to argument of counsel would be that the arguments would have to be based on introspection and fireside indications. There would be no sponsor for empirical studies about signs of forgery.[211]

If an FDE is allowed to testify, to any degree, the question becomes one of whether the FDE should be allowed to characterize the nature of the writing in ways that involve expert inferences (e.g., it appears to be "drawn" and shows signs of "hesitation") and to use expertise in teaching the jury about signs of genuine or simulated writing (e.g., unexplained tremor is a sign of slow writing, which is an indicator of tracing). And if the expert is allowed to make such assertions, then the jury will guess the expert's conclusion, so it probably does not make much difference whether the expert is allowed to take the next step of expressing an opinion about whether the signature is genuine or simulated. (In fact, it is possible that a gestalt ability to form this sort of opinion, and not the specific reasons for it, is what differentiates lay from expert performance on proficiency tests.)

There is a division of authority on how FDE expert testimony should be treated in signature authentication cases. It remains the general tendency of courts to issue global opinions allowing FDE testimony, regardless of the particular task.[212] But one case[213] admitted FDE

211. *See* sources cited *supra* note 209.

212. *See, e.g.*, United States v. Paul, 175 F.3d 906, 910–11, (11th Cir. 1999); United States v. Prime, 220 F. Supp. 2d 1203, 1205–06 (W.D. Wash. 2002), *aff'd*, 363 F.3d 1028 (9th Cir. 2004). These cases presented tasks other than the harrowing task of signature authentication scrutinized in this Article; however, they are examples of global or near-global endorsement and FDE expertise in terms that would include signature authentication expertise. In his presentation for this symposium, Professor Risinger noted that during the past four years, trial judges have shown a marked proclivity to admit

testimony that a signature was not genuine only with limits, and another case[214] excluded it entirely.

I think that FDE testimony on signature authentication should be admitted, subject to strict procedural limits. The first consideration is the nature of the experience-based expertise. FDEs are not harbor pilots, but they are not astrologers either. There is no sure "feedback loop," though in particular instances document examiners may learn that they are wrong and suffer a penalty.[215] Despite the absence of a regular feedback loop, it seems reasonable to believe that FDEs would learn from experience—their own experience, the experience of those under whom they apprenticed, the experience of the authorities in the field. One of the purposes for which the CTS tests are avowedly created is training, so they are one feedback loop.[216] The takers of the tests get specific feedback on the correctness of their answers, and the purchasers of the test materials can keep them and use them in teaching future apprentices. Second, the fact that FDEs are exposed to extraneous case information has a double effect. It diminishes the independent value of their expert judgment in a particular case, because their decision that a signature is genuine or not is likely influenced by information that has nothing to do with the signature. But in the long run the extraneous information gives them a basis for knowing what forgeries look like. Just the fact that they compare exemplars of unquestioned authenticity with questioned documents would be some basis for drawing inferences, since at least some of the questioned documents are sure to be simulations, so the fact that features appear with greater frequency in questioned documents than in the unquestioned standards is some evidence that those features are signs of simulation. When we add to that the extraneous information—for example, the document examiner knows

---

FDE testimony after string citing appellate court decisions that upheld admission of the testimony as within trial court discretion (and treating those appellate opinions as if they were ones that mandated the admission of FDE testimony). He concluded that the battle to exclude FDE testimony had been lost. Michael Risinger, Professor, Presentation at Hastings Law Journal Symposium: "Faces of Forensics: Identification and Behavior" (Mar. 21, 2008).

213. United States v. Starzecpyzel, 880 F. Supp. 1027, 1042 (S.D.N.Y. 1995).

214. United States v. Brewer, No. 01CR892, 2002 WL 596365, at *6–8 (N.D. Ill. Apr. 16, 2002). The document examiner would have testified that a signature was a traced forgery, the model for which was one of the exemplars that had been located. *Id.* at *6–7. The court held that the expert testimony was inadmissible. *Id.* at *8 (citing United States v. Saelee, 162 F. Supp. 2d 1077 (D. Alaska 2000); United States v. Fujii, 152 F. Supp. 2d 939 (D. Ill. 2000); United States v. Hines, 55 F. Supp. 2d. 62, 67–68 (D. Mass. 1999)).

215. Examples of exposure include the *Dreyfus* trial, *see* D.H. Kaye, *Revisiting* Dreyfus: *A More Complete Account of Trial by Mathematics*, 91 MINN. L. REV. 825 (2007), and the Clifford Irving/Howard Hughes hoax, *see* NICKELL, *supra* note 205. Identification of the signature on the Bush National Guard records also raised dangers of being proven wrong. *See* Howard Kurtz, *Rather Defends CBS over Memos on Bush*, WASH. POST, Sept. 11, 2004, at A07.

216. *See, e.g.*, CTS TEST No. 01-524, *supra* note 99, at 1 (containing standard statement that test samples may, in the option of the receiving labs, be used in a "training exercise").

that the other evidence indicates conclusively that the document is a forgery—then we can see how the examiner, and his predecessors and tutors, might develop a degree of justified confidence about what is suspicious and what is not. In other words, they have the opportunity to compare signatures known to be genuine with each other, and sometimes with signatures almost certain to be forgeries.[217]

In FDE expertise, we have a field that comes from an empirical tradition broadly defined, not a magical or superstitious one, like astrology. Its suppositions about the signs of forgery are plausible and logically consistent. It is one that has been endorsed by authorities that we ordinarily respect.[218] It is expertise that has not been discredited by the scientific method.[219] Its suppositions about signs of forgery are supported to some extent by systematic empirical research.[220] And finally, there are studies comparing lay and expert performance that indicate superior performance by experts under test conditions.[221]

We might wish for more. It would be nice, for example, to have studies comparing how mock juries do at the signature authentication task under the tutelage of an expert, compared to how they do on their own. Another condition could pit an anti-expert against an FDE expert, testing Professor Berger's hypothesis that this procedure might provide better protection for defendants than leaving the signature comparisons to untutored jurors who are not aware of the dangers of error.[222] Nonetheless, we do now have respectable studies that show superior expert performance. Those same studies can be used to show jurors something about expert error. The jury is certainly entitled to know, not only that experts have performed better than lay persons in avoiding the error of saying that a genuine signature is simulated, but also that even the experts have a significant error rate.

What about the loss of research if the evidence becomes admissible? It is quite possible that exclusion of FDE testimony would encourage more research. But it would be naive to assume that judges could force all forms of forensic science to rise to the level of DNA evidence, and it is not clear whether it would be wise to aspire to that level. Exclusion of

217. *See* sources cited *supra* notes 213–14.

218. For example, Dean Wigmore wrote the introduction to Osborn's first edition. *See* OSBORN, 1910 edition, *supra* note 28, at vii–ix. Also, many courts have endorsed the expertise. *See also* sources cited *supra* note 212.

219. *See generally* Black et al., *supra* note 14.

220. *See id.*

221. *See supra* Part III.

222. *See* Berger, *supra* note 123, at 1138.

> Is the defendant better protected when jurors during deliberation compare the uncrossed t's or undotted i's in the crime scene samples and the provided specimens, as they have probably seen done on some TV show, than if the court allows a prosecution expert to testify who is then demolished by Professor Saks?

Id.

evidence is a blunt tool with which to try to influence the allocation of resources toward research on forensic science. It cannot assess the value of competing claims to resources. A *Daubert* hearing is not the best forum to assess competing claims of forensic and medical research, or even the competing claims of different types of forensic research. Moreover, the amount of research needed would be enormous. The testing of basic factual assumptions requires an extensive research program, as is suggested by a careful listing of the basic premises of the field by two legal experts who support testing of those premises.[223] And the list is not complete. In the area of signature authentication, one can think of many other candidates for research—for example, whether simulations are marked by detectable tremor, blunt endings, patching, ink pooling, unnatural line width, etc. It would take an ambitious program to test each of these assumptions. Of course, more could be done on testing proficiency in the area that has been the particular focus of this Article—signature authentication—but even there, one must make resource allocation decisions. It may be more reasonable to move on to comparing how experts and lay subjects do in performing other handwriting identification tasks.

Finally, it is not completely clear that admitting identification evidence will always have the effect of dampening incentives for systematic research. In the eyewitness identification area, research has flourished despite the fact that eyewitness identification testimony is generally admitted.[224] One can think of plausible arguments for stricter exclusion of eyewitness identification testimony (for example, when the identification is not corroborated by other evidence, or when it is the result of a nonblind simultaneous lineup, or when the fillers do not sufficiently resemble the victim's description of the perpetrator). Stricter exclusion might be worthwhile on its own merits, but it is unclear what the effect of that approach would have on research. It is possible that admission of the evidence has stimulated research. The same thing may be said about confessions, where the research has generally come from doubters of the value of the evidence, not from proponents of confession evidence.[225]

Of course, there are many differences between eyewitness identification research and handwriting identification research, one of them being that issues of eyewitness identification overlap with academic psychologists' interests in the study of perception and memory by

223. Risinger & Saks, *supra* note 47, at 67–74.

224. *See* Roger C. Park & Michael J. Saks, *Evidence Scholarship Reconsidered: Results of the Interdisciplinary Turn*, 47 B.C. L. REV. 949, 960–64 (2006).

225. *See, e.g.*, RICHARD A. LEO, POLICE INTERROGATION AND AMERICAN JUSTICE (2008); Richard A. Leo et al., *Bringing Reliability Back in: False Confessions and Legal Safeguards in the Twenty-First Century*, 2006 WIS. L. REV. 479.

psychologists. But the fact that experts who doubt the value of the identification are allowed to testify has probably also stimulated research. If, as advocated in this Article, expert doubters are allowed to testify in response to FDE experts, that might also stimulate research by the doubters. We can also hope that the amount of academic interest will increase, for example by scholars interested in statistics and signal detection theory.[226]

My conclusion—admittedly a debatable one—is that the FDE experts should be allowed to testify in signature authentication cases. But that does not mean that they should have the unfettered freedom that they have sometimes been given in the past. Judge McKenna was correct, in the *Starzecpyzel* case, to take measures that would prevent FDEs from being confused with research scientists.[227] Procedural safeguards should include instructions that the expertise is not scientific, restrictions on the use of scientific jargon or science-invoking words such as "laboratory," and full discovery of detailed conclusions. Judges should consider enforcing discovery and reducing the role of personality by allowing the opponent to opt for admission of the written report in lieu of live testimony. When FDEs testify, witnesses with knowledge of the scientific method should be allowed to testify on the other side, questioning scientific status, proficiency, or the validity of conclusions. Moreover, in cases in which the stakes are high and the authentication question is crucial, the opponent of the expertise should be able to demand blind retesting by a neutral FDE.

226. David Faigman, Professor, Presentation at Hastings Law Journal Symposium: "Faces of Forensics: Identification and Behavior" (Mar. 21, 2008); *see also* David L Faigman, *Anecdotal Forensics, Phrenology, and Other Abject Lessons from the History of Science*, 59 HASTINGS L. J. 979 (2008).

227. *See supra* note 197 and accompanying text.

APPENDIX I: EXAMPLE TESTIMONY ABOUT A TRACED SIMULATION[228]

Q. You have stated that, in your opinion, the signature of the decedent on the purported will dated _____ is not genuine. Would it be helpful to use your exhibits in explaining how you arrived at your conclusion?

A. Yes, it would. While the signature in question, shown at the top of the chart, appears to be pictorially similar to the standard writings that are shown here [witness indicates], the writing in question takes on a dead, flat appearance when compared to the genuine counterparts. There is a significant difference in the line quality between the standard and questioned writings. There are many areas exhibiting tremor. Note the tremulous or shaky movement indicated by the arrows at the t -bar crossing in the first letter, and in the capital letters O and C. The writer of the standards moves his hand swiftly; the questioned signature is slowly written. The standard writings show tapered or feathered endings, whereas the questioned signature has blunt ending strokes. In addition, the variation of pen pressure differs significantly between the standard writings and the questioned signature. The document in question lacks the pen pressure variation seen in the genuine writings. This commonly occurs as a result of the copying process necessary in tracings or simulated writings.

Q. How is a tracing recognized?

A. Tracings may be recognized in several ways. If pre-existing writing has been erased, one would find pencil lead embedded in the paper fibers. If a carbon outline of the model writing was used, carbon traces are usually left on the document.

Q. Any other ways?

A. Yes. A tracing may also have the model signature indented onto the page by a stylus or other blunt instrument. Of course, the forger may simply use back light of some sort to copy a signature, such as by holding the paper up to a window and tracing the genuine writing. In such a case, the telltale indications of the tracing are not as prominent.

Q. Did you find any carbon tracings on the purported will?

A. No.

Q. Did you find any pencil markings or erasures?

A. No.

Q. Did you find any stylus markings or indentations?

A. No.

---

228. This appendix has been excerpted, with permission, from *American Jurisprudence Proof of Facts 3d.* 27 AM. JUR. 3D *Proof of Facts* § 108.

Q. How, then, do you know that the questioned signature is a tracing?

A. At first, I was unsure whether I had a tracing or a simulation, because there were no traces of pencil lead, carbon, or indentations. However, when the cognovit note was brought to my laboratory on _____[date], I knew the signature on the will was a tracing.

Q. Again, how do you know?

A. I had before me two signatures that were, for all practical purposes, the same. One of the basic rules of handwriting is that no person repeats handwriting with exact precision. Therefore, two writings that are exactly the same cannot both be genuine.

Q. If they are both the same, how can you tell which is the genuine writing and which is the tracing?

A. In this case, the signature on the cognovit note is written with greater speed, as evidenced in the initial "tick" strokes right here [witness indicates] and in the feathered ending strokes. The signature in question, however, has all the earmarks of nongenuineness, including slow, drawn movement, blunt ending strokes, and hesitations in the writing. My other chart, Plaintiff's Exhibit _____, shows this.

Q. Please continue.

A. I have photographed the signatures on both the cognovit note and the will with a transparent grid overlay. One can see the sameness of the beginning and ending positions of the signatures, and how the signatures follow the same grid alignment. Again, though, please note the differences in writing quality between these writings.

Q. Were you able to identify who traced the signature?

A. No. In cases of this kind, it is not possible to identify the writer of the tracing. The tracing is nothing more than a slow drawing of the model signature, and therefore lacks identifiable handwriting features.

APPENDIX 2: EXAMPLE TESTIMONY ABOUT FREEHAND SIMULATION[229]

Q. I'm going to ask you to explain how you arrived at your conclusion about the nongenuineness of the signature on the will. Would it be helpful for you to use your exhibits in explaining your findings to the jury?

A. Yes, very much so.

Q. Then I ask, with the court's permission, that you step down and, using the exhibits, explain how you arrived at your opinion.

A. Certainly. Now, the chart marked Plaintiff's Exhibit _____ shows signatures from both the questioned writing and the standards. Perhaps the first thing one notices in these signatures is the similarity between the standard and questioned writings. Pictorially, the standard and questioned writings look alike in letter design and shape. However, close examination shows significant differences between the bodies of writing.

Q. What differences are those?

A. There is a difference in the rhythm of writing as between the standards and the questioned signature. There is also a significant difference in the size of the writings. The questioned signature is smaller than the standard writings, even though there was approximately the same amount of writing space available. In addition, the standards are written with greater speed than the questioned signature. Slower writing is often found in a nongenuine signature, because the writer, being unfamiliar with how the letters are formed, must take greater care in the writing process and cannot execute the writing freely.

Q. Any other differences?

A. Yes. The letter heights and ratios found in the bodies of writing are significantly different. The standard writings are narrow, oval, and more angular in form than the questioned signature forms, which are more compressed. I might also point out that the author of the standard writings places the first and last names close together, while the writer of the questioned signature allows for more space between the names. In addition, there is a significant difference in the forms of letters and letter connections as between the standard and questioned writings.

Q. Please explain.

A. The letter l in the standards is written with a vertical or rightward slant, while the l in the questioned signature angles to the left. The letter i in the questioned writing differs in that it is extremely short

229. This appendix has been excerpted, with permission, from *American Jurisprudence Proof of Facts 3d*. 27 AM. JUR. 3D *Proof of Facts* § 110.

when compared to the i in the standard writings. The i -dot in the questioned writing is made with a rightward dash, unlike the i -dot in the standard writings. The letter e in the standards is narrow and vertical, while in the disputed signature the e is short and squatty. The capital letter T in the last name is typically crossed with a long, sweeping stroke in the standard writings, but is crossed with a much shorter stroke in the questioned document. This is different from any of the standards I examined. The letter h differs also. In the standard writings, the letter is constructed with a slightly hooked downstroke, followed by a second stroke of the letter, forming a definite v-shaped wedge coming from the base of the h. In the questioned writing, the letter h has quite a different appearance.

Q. Are there any other differences in letter form?

A. There are several other such differences, yes. There is a difference between the standard and questioned writings in the rhythm and design of the letter m, for example. Again, we see compression in the questioned writing that is not found in the standards. Although the letter p begins in a similar fashion in both the standard and the questioned writings, they are different. This is best seen at the point where the two portions of the lower loop of the p intersect. In the standards, the pen moves diagonally, but in the questioned signature, this portion of the letter moves horizontally. The questioned p is definitely flatter and proportionally shorter than in the standard writings. Moreover, the letters s are indicative of two different writers. The letters in the standards display only a hint of a hook at the beginning of the s, whereas the questioned signature shows a wider stroke. The ending formation of the letter s is not the same. In the standards, the bottom of the letter is very angular on the left side, while the questioned s is rounded.

Q. Please continue.

A. The second o is also formed differently as between the standard and questioned writings. Again, you can see the oval form of the standards, in contrast to the rounded form found in the questioned signature. In the standards, the letter n is nondescript and, in fact, illegible. The questioned n is more defined—not a significant difference in itself, as a person can write with greater definition from one moment to the next, but nevertheless a difference in the flow of the final stroke. The standards writer consistently ends with a blunt stroke, unlike the writer of the questioned signature, which ends with a sweeping flourish.

Q. You also mentioned some differences you observed in letter connections?

A. Yes. In the standards, the h is continued with a downstroke of the pen, and at this point, the writer moves on to form the letter o. In the questioned writing, the h begins in a similar fashion, but that is where the

similarity ends. Unlike the writer of the standards, the writer of the questioned signature retraces part of the stem and then moves the pen to the right in a sweeping horizontal direction. There is a pen-lift between the h and the o that is not found in any of the standard writings. Also, there is a fundamental difference between the standard and questioned writings in the construction of the first o in the last name. In the standard writings, the letter o initiates from the last stroke of the preceding h. The pen then moves upward, flowing in a counterclockwise motion. Note how the o is flat on the left side but curved on the right side. The writer of the standards regularly ends the letter on the left side and connects the "om" combination, where the m begins from the top left side of the preceding o.

Q. And that differs from what you observed in the questioned writing?

A. Yes. The first letter o in the questioned signature begins at the top of the letter, moves in a counterclockwise direction, and ends on the right side of the letter, connecting with the letter m from the right side. That is another indication of two writers rather than just one.

Q. So your opinion, once again, is what?

A. After closely comparing the known signatures of the decedent with the questioned signature, one can see that, although the writings bear some similarities, the disputed signature is clearly a simulation of the true writing, and not a genuine signature.

APPENDIX 3: 1988 CTS TEST RESULTS[230]

PART 1. 1988 CTS TEST: TASKS TESTED

In the text, I extracted the signature authentication tasks from the 1988 CTS test. It also tested other tasks, which will be described here. After that description, the reader will find a detailed tabulation of the signature authentication tasks.

In the test scenario, the six signed receipts were the questioned documents, designated as Q1, Q2, Q3, Q4, Q5, and Q6. The request exemplars of known signatures were designated K1, K2, K3, K4, and K5. The givers of the exemplars signed their own names and all of the other names in the questioned documents repeatedly.

The receipt Q1 was signed with the name. "Sharon D. Clayborne." It was actually signed by Richard D. Osbourn, the "driver" whose exemplars were designated K5, not by the real Sharon D. Clayborne, whose exemplars were designated K1.

The receipt Q2 was signed with the name "Lisa D. Bridgeforth." t was actually signed by Lisa D. Bridgeforth, whose exemplars were designated K2.

The receipt Q3 was signed with the name "Cynthia Y. Boone." It was actually signed by Richard D. Osbourn, the "driver" whose exemplars were designated K5, not by the real Cynthia Y. Boone, whose exemplars were designated K3.

The receipt Q4 was signed with the name "Joanna Neuman." It was actually signed by the real Joanna Neuman, whose exemplars were designated K4.

The receipt Q5 was signed with the name "Linda M. Ninestine." There were no exemplars from the real Linda M. Ninestine. The signature on receipt Q5 was not signed by any of the subjects from whom exemplars had been obtained.

The receipt Q6 was signed with the name "Linda D. Wentworth." It was actually signed by Richard D. Osbourn, the "driver" from whom the exemplars designed K5 had been obtained. There were no exemplars from the real Linda D. Wentworth.

The test takers were asked to compare each of the questioned signatures (Q1 through Q6) with each of the known handwriting samples (K1 through K5) and to identify or exclude the writer of the exemplar as the writer of the questioned signature. In identifying or excluding, the test takers were given the choice of a firm conclusion (the questioned signature "was" or "was not" written by the author of one of the sets of exemplars) or a qualified one (the questioned signature "was probably"

230. CTS REPORT NO. 88-5, *supra* note 55; *see also supra* note 93 and accompanying text.

or "was probably not" written by the author of one of the sets of exemplars). Test takers were also given the choice of "no conclusion" or "other (specify)."

A completely correct answer on the test would have been:

1. Richard Osbourn was the author of Q1, Q3, and Q6.
2. Lisa Bridgeforth was the author of Q2.
3. Joanna Neuman was the author of Q4.
4. None of the exemplar givers was the author of Q5.

PART 2. TABULATION OF SIGNATURE AUTHENTICATION RESPONSES ON 1988 CTS PROFICIENCY TEST[231]

Receipt Q1:

- Correct answer: The questioned document is not the authentic signature of Sharon D. Clayborne.
- Results: 100% correct (48/48) (all respondents gave affirmatively correct answers)

Receipt Q2:

- Correct answer: The questioned document is the authentic signature of Lisa D. Bridgeforth.

Results:

- All opinions: 87.5% correct (42/48) (42 correct, 1 wrong, 2 inconclusive, 3 "other")
- Called opinions: 97.6% correct (42/43) (discarding inconclusive/other responses)

The "other" comments were:

- "definite similarities"
- "inconclusive with investigative leads for collection of additional evidence (standard writing) or notation . . . [that] case may not be resolved via F.D.E. Examination."
- "Significant similarities noted"

231. In tabulating the answers, I have excluded the answers of lab 517, which lodged a blanket protest against every question on the test, writing that "Due to the quality of the photographs and the type of known handwriting specimens submitted, a definite conclusion could not be reached . . . ." I think it is more appropriate to treat this statement as a nonresponse than as an "inconclusive." CTS REPORT No. 88-5, *supra* note 55, at 24.

Thus, two of the "other" comments lean toward the correct answer, and one calls for additional exemplars.

Receipt Q3:

- Correct answer: The questioned document is not the authentic signature of Cynthia Y. Boone.
- Results: 97.9% correct (47/48) (all respondents gave affirmatively correct or wrong answers)

Receipt Q4:

- Correct answer: The questioned document is the authentic signature of Joanna Neuman. (The authentic Neuman signature posed the biggest problems for the FDEs. Some of the written comments suggest that the reason might be that Neuman's exemplars had a good deal of intrawriter variation and that her signature did not have many distinguishing features. It is also possible that Neuman made some attempt to vary or disguise her hand when she signed the "receipt," though the released test description gives no reason for thinking that this occurred.)

Results:

- All opinions: 52.1% correct (25/48) (25 correct, 3 wrong, 11 inconclusive, 9 other)
- Called opinions: 89.2% correct (25/28) (discarding inconclusive/other responses)

The "other" comments were:

- "possibly the writer of"
- "no association"
- "some indications, want more known"
- "some similarity and could be"
- "there were enough similarities noted to warrant the examination of additional known standards"
- "should not be eliminated"
- "add'l exemplar and analysis required"
- "cannot be identified or eliminated"
- "definite similarities"

Classifying these comments, seven show at least a suspicion that Neuman was the writer (a correct answer), one found "no association" with any of the exemplars from any of the subjects (a wrong answer), and one simply called for additional exemplars and analysis.

Overall tabulation of called opinions:

- False authentication error: 1% (1/96)
- False elimination error: 6% (4/71)

***